

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

## Science in China

Money, ambition and innovation in the research superpower **PAGE 451**



### ECONOMICS

#### NATIONAL WEALTH

The search for something more useful than GDP

**PAGE 472**

### ASTROPHYSICS

#### ASTRONOMY'S NEW WAVE

The first gravitational-wave source modelled

**PAGES 470 & 512**

### PARTICLE PHYSICS

#### EXPERIMENTS IN THEORY

A quantum optics simulation of fundamental physics

**PAGES 400 & 516**

**NATURE.COM/NATURE**

23 June 2016 £10

Vol. 534, No. 7608



# THIS WEEK



## EDITORIALS

**GUNS** New hope for firearms research in United States **p.436**

**WORLD VIEW** Discuss climate change in the present tense **p.437**

**GOBBY GLUE** Sticky saliva the secret to terrific tongues **p.438**

## Way of the dragon

*To boost its research quality and innovation, China must strengthen its scientific foundations and let researchers — not policymakers — set the agenda for innovation and discovery.*

There is increasing excitement over China's scientific rise. The nation has more researchers than any other country and it is rapidly catching up with the United States in the number of scientific papers published. But there are lingering questions — both within China and outside — about the quality and inventiveness of science coming out of the country.

Concerns over science in China go to the very top. Xi Jinping, China's leader, offered a particularly harsh assessment late last month at a meeting of the country's leading scientific academies. He went so far as to say "the country's S&T foundation remains weak".

Xi has a point. Many of the inventions that gave rise to some of the most important scientific work in China — CRISPR-Cas9 gene-editing tools among them — are the products of colleagues overseas. Xi put it like this: "The situation, in which our country is under others' control in core technologies of key fields, has not changed fundamentally."

From that angle, China still looks like a nation of large-scale implementers. Take an idea, especially one that requires scale, and China is there to jump on it. That is not a bad place to be — the genome-sequencing giant BGI and a new generation of sequencing rivals are a clear sign of just how productive scale can be. But that is application, not the kind of breakthrough that Xi seeks.

That's why the country's first scientific Nobel prize, awarded last October to Tu Youyou for her role in developing the antimalarial drug artemisinin, provoked pride but also soul-searching. It was a discovery from a bygone era, not a product of the current research structure — and many wonder whether today's system will yield any big discoveries.

In a special issue this week, *Nature* looks at China's potential and the obstacles it faces (see [www.nature.com/chinafocus](http://www.nature.com/chinafocus)). Xi told the meeting that "scientists should be allowed to freely explore and test the bold hypotheses they put forward". He encouraged the development of a system in which science policy is created by scientists, rather than at the whim of officials, and alluded to experts who "should no longer have to follow their superiors' orders".

If anyone can break the bureaucrats' hold on scientific policy-making, it is Xi, who has emerged as China's strongest leader in decades. He has already taken on, and taken down, numerous political foes. And yet, as China implements its latest five-year plan and overhauls its major funding mechanisms, there is reason to wonder how much things will change.

Xi couches much of his support for science as the quest for translatable results. Scientists should, he says, solve urgent economic and industrial problems. Support for technology firms is a high priority. These are fine

objectives, but they suggest continued top-down policymaking. The balance between encouraging basic research and demanding technological output must be guarded closely, or scientists will be pressured to do only translatable research and China will tread on the freedom of scientific pursuit that Xi holds is essential.

Although Xi seems to understand the scientific thirst for independence and freedom, the ongoing question is whether China will offer that. This includes freedom to use tools such as Google Scholar.

**"Truly pioneering science is to be cultivated, not commandeered."**

Xi faces some of the greatest battles of China's recent past: military tussles in the South China Sea have raised the political stakes abroad, economists talk of a dangerous slowdown, and environmental problems are frustrating citizens at home and threatening the country's international stature. Xi vows to raise spending on science, but it would be a mistake to think that increasing spending on research and development will solve all the issues of the homeland, make food and drugs safe, resolve the problem of an ageing population and get rid of the disparities between urban and rural China.

At the meeting, Xi said: "Currently, the state needs the strategic support of science and technology more urgently than any other time in the past." But truly pioneering science is to be cultivated, not commandeered. How well that distinction is maintained will determine much of what lies ahead. ■

## Data sharing

*Pooling clinical details helps doctors to diagnose rare diseases — but more sharing is needed.*

When doctors in Ottawa saw a child with an unusual developmental disorder last year, they were stumped. Their patient had an abnormally small head and face and had been slow to develop. They sequenced the child's genome hoping to find a genetic explanation, but came up with too many possible candidate genes to pinpoint a likely culprit. This still happens a lot in medicine: people with rare problems go undiagnosed. And that's one reason behind a big push in science in recent years — the pooling and sharing of clinically relevant information.

In the Ottawa case, the doctors got lucky. They were able to search a database that contained information about other patients with undiagnosed diseases, and when they did so they found a second person with similar symptoms — and an identical mutation in one



**SCIENCE IN CHINA**  
A Nature collection  
[nature.com/chinafocus](http://nature.com/chinafocus)

gene, *EFTUD2*. The finding allowed the Ottawa doctors to diagnose their patient with a disease called mandibulofacial dysostosis with microcephaly, and to begin to understand why mutations in *EFTUD2* cause the disease's symptoms.

That's the upside of the new era of data sharing. But there is a possible downside too: invasion of privacy. Massive genetic studies in countries such as the United States, Qatar, Saudi Arabia and Brazil are collecting genetic data on millions of people, so there is a chance that a person's identity could be dragged from those data — especially if they are linked to clinical information, such as medical history. The risk is that someone who volunteers their DNA could see their medical problems opened to public scrutiny.

This is a legitimate concern for many researchers, and is one reason why data sharing is easier said than done. Others include the lingering sense of ownership, and the career benefits offered to those who have privileged access. Those concerns relate to the standard model of data sharing, in which different groups of scientists deposit their results into centralized databases. This model has had some success, but researchers have already encountered problems, such as how to grant and control access to the pooled information.

Pooling it in the first place becomes more difficult as the data sets get larger and the underlying techniques more varied. Imagine the difficulty of finding a specific book by gathering all the contents of a dozen different national libraries and then devising a way to integrate the numerous ways in which they are filed, tracked, recorded and made available. It would be much easier to ask each library whether it holds that book. What if data sharing in science could go the same way?

The diagnosis of the Ottawa child shows that it can. The doctors tapped into a system that is part of the Matchmaker Exchange, which allows researchers to query multiple databases of information on patients with undiagnosed rare diseases. A doctor can feed the system information about a patient's symptoms and genetic make-up, and then ask it whether other people have them too. (Normally, it's hard

for doctors to find other patients with similar rare diseases; often they learn about such cases by word of mouth.)

The Matchmaker Exchange exemplifies a subtle shift in how researchers think about data sharing — and one that more scientists should engage with. It was created by the Global Alliance for Genomics and Health, a 3-year-old organization with more than 700 members from 70 countries that aims to help researchers,

***“As technology to permit targeted data access improves, so will smart sharing.”***

doctors and patients to make scientific progress by sharing data (see *Global Alliance for Genomics and Health Science* 352, 1278–1280; 2016).

The alliance is creating technological tools that allow researchers to find out where data that are relevant to their patients are held around the world. It aims to make data not just shareable but discoverable, too. Doing this allows those who produce the data to keep more control of the information. It also streamlines searches. For example, researchers looking for a diagnosis want to know the symptoms that other doctors have seen in people with particular genetic traits. Thus they just want to know who might have seen these mutations and what symptoms might have been observed in patients who have them; they don't want to comb through all the existing databases of genetic information themselves.

Of course, there are still many instances in which accumulating and sharing large amounts of data — on particular genetic traits, for example — is essential and valuable. The gene-testing company Myriad Genetics is locked in a tussle with doctors and patients who want it to open up its massive database of information on variations in the *BRCA1* and *BRCA2* genes, which are linked to a higher risk of breast and ovarian cancer. (Another alliance project, the BRCA Exchange, seeks to provide easily searchable interpretations of BRCA variants that have been shared by groups outside Myriad.)

But in other cases, data access works best, for both sides, when the requests for information are targeted at specific traits. And as the technology to permit that improves, so will smart sharing. ■

# At gunpoint

*The problem of gun violence in the United States must finally be addressed.*

**I**t has been a bloody year in the United States. So far, the country has lost around 6,000 lives to gun violence — dozens of them in mass shootings in public spaces. The attack that left 49 men and women dead in Orlando, Florida, this month is, by some counts, the 136th mass shooting in the United States just this year.

Mourning — and then moving on — in the wake of a mass shooting has become a sombre tradition. But after Orlando, a new development emerged. On 14 June, the American Medical Association (AMA) declared gun violence a public-health crisis, and announced that it will apply its considerable lobbying power to pressure Congress to fund research into this violence. It is cause for optimism that a lengthy freeze on federal funding for such research — particularly at the Centers for Disease Control and Prevention (CDC) — may soon thaw.

It makes sense that this push would come from the medical community: it has a front-row seat on the violence. “Here we are again,” physicians wrote in a *New England Journal of Medicine* editorial in January, following a shooting in San Bernardino, California, that killed 14 and injured 22. Six months later, at a press conference following the Orlando tragedy, one surgeon choked back tears as he described the chaos in an emergency room filled not only with the injured, but also with hundreds of their panicked friends and families. Another coolly

described the reality that surgeons at his Orlando trauma centre face daily: people wounded by high-calibre assault rifles, once considered to be the exclusive domain of the military, now flooding into civilian emergency rooms.

Yet while doctors struggle to treat the wounded, the CDC has been hamstrung in tackling fundamental public-health questions about the causes of gun violence and its possible solutions. An amendment placed on appropriations bills since 1996 has prohibited federally funded research from advocating gun control — a provision that some have interpreted as making gun-violence research broadly off limits.

In 2013, US President Barack Obama explicitly stated that such research should take place and need not be interpreted as advocacy, but Congress failed to allocate funds in the CDC budget to support it. (The US National Institutes of Health, which has more discretion in how it applies its funding, has sponsored some gun-violence research following Obama's announcement.)

The AMA is a lobbying powerhouse: in 2015, it was the fourth-largest lobbyist in the country. If it chooses to make gun-violence research a high priority, it has the resources to make headway. But it will take a tremendous push — and coordination with other stakeholder organizations — to do so.

In the wake of the Orlando shooting, lawmakers followed what has become a legislative post-mass-shooting tradition: the rapid-fire proposal — and equally rapid rejection — of bills intended to address the country's gun-violence crisis. Earlier this week, the US Senate defeated five such measures. Similar proposals, including one intended to explicitly allow research into gun violence, met the same fate last December. But with concerted effort from the AMA and others, perhaps the United States will break with these traditions. ■



## Bring climate change back from the future

The 'shock' over an Australian extinction shows that we still don't accept that global warming is a problem for now, says **James Watson**.

Climate change has claimed its first mammal casualty, with the reported extinction of the Bramble Cay melomys (*Melomys rubicola*). The last of these Australian marsupials is thought to have disappeared around 2009, but the release last week of a report by the Queensland government stating the probable extinction of the species and the cause — sea-level rise induced by climate change — made worldwide news.

The death of the last individual of the last population of a mammal species, indeed of any species, is as irreversible as it is profoundly sad. Yet the widespread coverage of this extinction and the subsequent outpouring of concern from across society tapped into something else. Species go extinct every day with little fanfare or report. The last Australian mammal to go extinct before the melomys was the Christmas Island pipistrelle (*Pipistrellus murrayi*) in 2009, with almost no press. The melomys extinction was covered because it ended the idea that climate change will be a concern for species only in the future. That reflects a fundamental, widespread problem with how we think about and report on climate change, especially when it comes to nature and conservation. Too many people still think that climate change is a problem that we can deal with later.

It's easy to see why. Climatologists use long-term forecasts, on timescales such as 50–100 years, and for good reason. It takes long periods of time for alterations in atmospheric concentrations of greenhouse gases to cause change. Looking ahead for a scientist brings increased certainty — we know that there will be a problem to address. And politicians like to emphasize the long term for the opposite reason: they can stress the uncertainties in the detail, and talk about action without needing to take any. Yet these distant forecasts have also become the basis of how people assess and communicate the probable effects of climate change on species and ecosystems. And as the Bramble Cay melomys shows, we are seeing those impacts now.

The world's climate system is already seriously disrupted: the global average temperature is already nearly 1 °C warmer than it should be. Across Earth, we are seeing radical shifts in daily temperatures, rainfall regimes and the timing of seasons, as well as overall increases in the number and intensity of droughts, cyclones and floods. It is now accepted that we have moved beyond the natural climate cycle and that, even if climate-mitigation policies are implemented immediately, it will take centuries to recover.

Nature is in the firing line. Climate change introduces new threats and speeds up existing declines. There is an avalanche of extinctions coming because of the direct impacts of change — temperature, rainfall and sea-level rise. But that is not the end of it. Climate change also interacts

with other major forces that have precipitated the current extinction crisis — most of which are also driven by human actions. Vulnerable human communities are responding to the changing climate, and adding significant pressure to already degraded ecosystems. For example, expansion of agricultural activities owing to more favourable rainfall regimes across the Albertine Rift and the valleys of the Congo Basin now increasingly threatens the most biodiverse regions in Africa.

If we are going to have a fighting chance to avert the current extinction crisis, we must accept and communicate that climate change is already upon us and that proactive action is needed now. We should not treat the news of the extinction of the melomys as an interesting question for Trivial Pursuit or an undergraduate exam — we need to treat it as a lesson.

This species did not live in a place where its existence came into conflict with other societal needs, such as good farming land or places to live. It was on an uninhabited island, effectively protected from other threats. A wide range of actions could have been taken to manage its population without causing conflict with other competing agendas.

Australian marsupials are well researched, and given the melomys's habitat requirements, the islands' low elevation and the fact that there is widespread knowledge of increasing sea levels across coastal Australia, it was not hard to work out that the species was in dire trouble. Yet almost nothing was done in time: there were no proactive plans to monitor the melomys, move a few individuals to create a rescue population or create a simple sea-level barrier. No action was

taken because of the attitude that climate change is not really happening yet, and there is time to sort it out.

This is unacceptable. We need a fundamental shift in how the scientific community, the media, policymakers and environmental funders view and discuss climate change. When we think about the impact of climate change on biodiversity, we need to start framing the issue as something that is already well under way and that, in conjunction with other threats, needs to be managed now. Crucial to this will be research on what species are immediately threatened by climate change, followed by plans to help them to survive. It will be complicated, but to give nature a chance, we need to harness the fears of the future to address the realities of the present. ■

**James Watson** is an associate professor at the University of Queensland in Brisbane, president of the Society for Conservation Biology and director of the science and research initiative at the Wildlife Conservation Society.  
e-mail: [jwatson@wcs.org](mailto:jwatson@wcs.org)

ACCEPT AND  
COMMUNICATE THAT  
CLIMATE CHANGE IS  
ALREADY  
UPON US  
AND PROACTIVE  
ACTION IS  
NEEDED NOW.

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## METABOLISM

### Bone hormone boosts exercise

A hormone released from bones enhances muscle function during exercise, giving old mice the capabilities of young ones.

Gerard Karsenty of Columbia University Medical Center in New York City and his colleagues found that blood concentrations of a hormone called osteocalcin increased during aerobic exercise in mice, monkeys and people. The hormone helped muscles to adapt to exercise by increasing their uptake and use of glucose and other nutrients. Mice that lacked the gene for osteocalcin had diminished exercise capacity.

Blood concentrations of osteocalcin declined as animals aged, and administering the hormone to 15-month-old mice gave them the exercise capacity of 3-month-old animals.  
**Cell Metab.** 23, 1078–1092 (2016)

## ASTRONOMY

### Early galaxy has wisps of oxygen

Astronomers have detected oxygen in a 13-billion-year-old galaxy — the first time that the gas has been found at such an early stage of the Universe.

A team led by Akio Inoue at Osaka Sangyo University in Daito, Japan, used the powerful Atacama Large Millimeter/submillimeter Array (ALMA) in Chile to measure the chemical make-up of the galaxy, which was discovered in 2012. Oxygen was only one-tenth as abundant as it is in the Sun, and the galaxy seemed to be

low in neutral gas and dust.

Such characteristics may have allowed ultraviolet light from the stars of this and other similar galaxies to escape and ionize the hydrogen atoms in the early Universe, eventually generating the levels of ions seen today.

**Science** <http://doi.org/bj5z> (2016)

## NEUROSCIENCE

### Smart birds have big brains

Birds that sing or use tools have about as many neurons in their brains as monkeys do.

Pavel Němec at Charles University in Prague and his

colleagues measured the brain size of birds from 28 species, and counted the number of cells in the organs. They found that intelligent birds such as parrots and songbirds have larger brains relative to their body size, with much higher neuron density, than do less-intelligent birds such as chickens. Moreover, a higher proportion of the neurons were located in the forebrain, which controls higher cognitive function.

Such high neuronal densities could be contributing to the intelligence of the birds, the authors suggest.

**Proc. Natl Acad. Sci. USA** <http://doi.org/bjzx> (2016)



## BIOPHYSICS

### Chameleons' sticky spit grabs prey

Adhesive mucus allows chameleons to snare insects with their long tongues.

Pascal Damman at the University of Mons in Belgium and his colleagues collected mucus from the tongue pads of veiled chameleons (*Chamaeleo calyptratus*; pictured) and found that it is 400 times more viscous than human saliva. Using a model of chameleon tongue strikes, the

team estimated that the mucus allows the animal to capture insects that are up to 60% of its body size — larger than its natural prey.

The size of prey a chameleon can nab is therefore not limited by the stickiness of its tongue, the authors say.

**Nature Phys.** <http://dx.doi.org/10.1038/nphys3795> (2016)

## NEUROSCIENCE

### Microbe makes mice social

Female mice that eat a high-fat diet produce litters with social deficits that are linked to changes in the offspring's gut bacteria.

Mauro Costa-Mattioli at Baylor College of Medicine in Houston, Texas, and his colleagues compared offspring from mothers that ate a high-fat diet with those from mothers on a normal diet. The high-fat-diet offspring spent less time interacting with other mice, and had reduced bacterial diversity in their

STEPHEN DALTON/NATUREPL.COM

MIKE DEMBECK

guts. In the animals' brains, researchers found fewer neurons containing oxytocin, a hormone that is related to social behaviour. Electrical signalling in the ventral tegmental area, a brain region that processes rewarding stimuli, did not strengthen as it normally does following new social interactions.

The team identified a gut bacterium, *Lactobacillus reuteri*, that reversed these abnormalities when it was given to the high-fat-diet offspring.  
**Cell** 165, 1762–1775 (2016)

## MEDICAL DEVICES

## Insect-eye camera peers inside gut

Mini cameras at the end of a probe that are designed to 'see' like an insect's compound eye could eventually be used in medical endoscopes.

Ömer Cogal and Yusuf Leblebici at the Swiss Federal Institute of Technology in Lausanne built a dome-shaped device measuring 10 millimetres wide with 24 tiny cameras covering its surface (**pictured**). In between the cameras are small, light-emitting fibre-optic cables. The cameras are placed in a way that provides a 180° × 180° or 360° × 90° panoramic field of view. The researchers found that the resolution of the system was 1,000 times higher than other insect-eye-inspired devices of a similar size.

When tested inside a tube mimicking a human colon, the device could reveal areas that are normally missed by conventional endoscopes.  
**IEEE Trans. Biomed. Circuits Syst.** <http://doi.org/bj3q> (2016)



IEEE 2016

## THERAPEUTICS

## Antibody double trouble for HIV

Genetically engineered human antibodies that bind to two targets on HIV could one day be used to treat and prevent the disease.

'Broadly neutralizing' antibodies can block various HIV strains, but the virus can overcome them by changing the viral protein that the antibodies recognize. To combat this viral escape, a team led by Jeffrey Ravetch at the Rockefeller University in New York City developed HIV antibodies that can recognize two different spots on the envelope protein that adorns the virus. One of these bispecific antibodies lowered viral levels in HIV-infected mice by more than tenfold in comparison with broadly neutralizing antibodies.

An independent team led by David Ho, also at the Rockefeller University, created bispecific antibodies that recognize both the HIV envelope protein and human proteins that HIV uses to infect immune cells. The most potent of these antibodies protected mice from becoming infected with HIV and decreased viral levels in those already infected.  
**Cell** 165, 1609–1620; 1621–1631 (2016)

## ORGANIC CHEMISTRY

## Plastic waste turned into fuel

Plastic from bottles and bags can be degraded into liquid fuels and waxes using available catalysts.

Polyethylene is the world's most common plastic, but is difficult to break down, typically requiring temperatures higher than 400 °C. A team led by Zheng Huang at the Shanghai Institute of Organic Chemistry in China and Zhibin Guan at the University of California, Irvine, used low-cost and widely available reagents



called light alkanes, along with key catalysts, to convert polyethylene into oils and waxes. The team broke down various forms of the plastic into oil at 175 °C within 4 days, at efficiencies ranging from 51% to 86%.

The team also used their process to convert 57–72% of everyday plastic waste such as bottles and bags into oils.  
**Sci. Adv.** 2, e1501591 (2016)

## ECOLOGY

## A climate refuge for trees

Forests in northeastern North America (**pictured**) could thrive in a warmer climate.

How trees will react to a warmer environment is unclear; low average temperatures hamper their growth but higher temperatures can limit water availability. Loïc D'Orangeville of the University of Quebec at Montreal in Canada and his colleagues used tree-ring data from more than 16,000 stands of black spruce (*Picea mariana*) across Quebec to track growth between 1960 and 2004. They found that north of a latitude of 49°N, increased temperature had positive effects on tree growth, despite the lower availability of water. Below that latitude, however, only an increase in water availability boosted tree growth.

Although boreal forests in central and western North America might be negatively affected by climate change, northeastern areas could act as

a refuge for certain trees, the authors suggest.

**Science** 352, 1452–1455 (2016)

## NEUROBIOLOGY

## Mitochondria make nerves grow

Enhancing the mobility of energy-producing structures called mitochondria in injured neurons helps these cells to regenerate in mice.

After an injury, some young neurons can regrow their long signalling arms known as axons, but mature cells cannot. Zu-Hang Sheng of the National Institutes of Health in Bethesda, Maryland, and his colleagues knocked out a gene called *Snph* in cultured mouse neurons. The gene encodes a protein, syntaphilin, that anchors mitochondria inside cells. The team found that 69% of young neurons lacking the gene began to form growing tips, compared with only 44% that had the gene. Similar effects were seen in adult mice, and injured mature neurons showed an increasing ability to regenerate with declining levels of syntaphilin.

Replenishing the energy supply in damaged axons by boosting the transport of mitochondria could be one way to treat nerve injuries, the authors suggest.

**J. Cell Biol.** <http://doi.org/bj3n> (2016)

## NATURE.COM

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

# SEVEN DAYS

The news in brief

## EVENTS

### Polar evacuation

The US National Science Foundation has launched a rare and daring mid-winter effort to evacuate a crew member from its South Pole research station. The person's name and medical condition have not been released, but 48 people are spending the winter season at the Amundsen-Scott South Pole Station. On 20 June, two Twin Otter aeroplanes arrived at the British Rothera station on the Antarctic Peninsula, which is being used as a staging base for the mission. The rescue team is awaiting suitable weather to make the 2,400-kilometre flight from Rothera to the South Pole station.

### Scientists say stay

As the United Kingdom prepares to vote on whether or not to stay in the European Union, 5,000 researchers have signed a letter to *The Times* newspaper warning that a British exit, or 'Brexit', will damage science. The letter, organized by the anti-exit group Scientists for EU and published on 20 June, is the latest intervention by researchers ahead of the 23 June referendum. Many scientists say that leaving the EU would be bad for research, but there is also a vocal contingent that holds the opposing view.

## POLICY

### Diluted vaccine

Yellow-fever vaccine could be used effectively at a diluted dose should the ongoing epidemic in Africa worsen, the World Health Organization's expert committee on immunization agreed on 17 June. The epidemic — the worst in almost 30 years — has infected a reported 3,137 people and



BILL INGALLS/NASA VIA GETTY

## Space-station crew comes home

An International Space Station crew — Tim Peake of the European Space Agency (ESA), Yuri Malenchenko of Roscosmos and NASA's Tim Kopra (pictured, left to right) — returned to Earth in a Soyuz capsule on 18 June, landing in the plains of Kazakhstan. Peake is the first

British ESA astronaut, and Malenchenko has now spent 828 days in space, the second most of anyone. After arriving in orbit last December, the three took part in astronaut-health studies to investigate how our eyes, brains and immune systems adapt to long-duration space flight.

killed 345 in Angola, the worst-affected country. The full-strength vaccine conveys lifelong protection, but mass immunization to control the epidemic has depleted stockpiles. Using the vaccine at one-fifth of its normal strength would provide protection for at least 12 months — adequate in an emergency situation — and would leave five times as much vaccine available.

### JAXA pay cuts

Three top executives of the Japan Aerospace Exploration Agency (JAXA) are taking a temporary pay cut because of the loss of the agency's Hitomi X-ray astronomy satellite. JAXA president Naoki Okumura and two others will reduce their pay by 10% for four months from July, the agency announced on 15 June.

After a successful launch in February, Hitomi was lost in March after engineers sent it the wrong software command, causing it to tumble out of control and break apart.

## RESEARCH

### Space fire

NASA scientists ignited the largest-ever deliberate fire in space on 14 June. Engineers used a remote-controlled hot wire to set light to a sheet of cotton and fibreglass measuring 1 metre by 0.4 metres. The material was housed in an uncrewed Orbital ATK Cygnus cargo ship that had just left the International Space Station. Instruments placed around the sheet monitored the fire as it burned for around eight minutes, and sent data back to Earth. The

Space Fire Experiment, or Saffire, is the first of three tests designed to understand how fire spreads in microgravity, and ultimately to improve fire safety for astronauts.

### Coffee and cancer

Coffee is unlikely to cause cancer, said the World Health Organization's cancer agency on 15 June, reversing its previous guidance. In 1991, the International Agency for Research on Cancer in Lyons, France, had described coffee as "possibly carcinogenic to humans". The latest guidance, which is based on a review of more than 1,000 studies, made no such connection. However, beverages consumed at a higher temperature (above 65 °C), such as piping hot tea and maté were deemed "probably carcinogenic" on

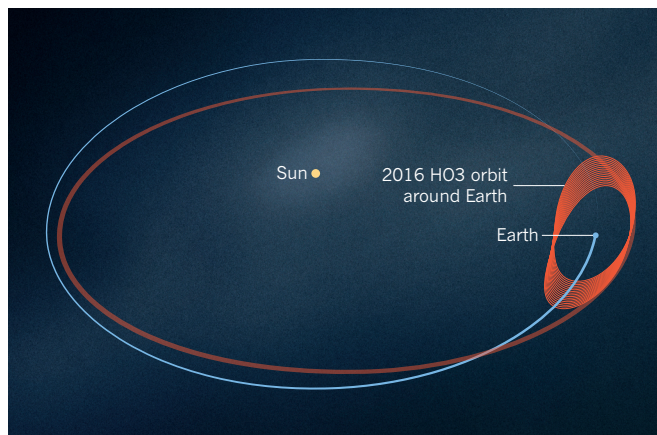
NASA the basis of epidemiological evidence. The drinks had a particular link to oesophageal cancers (D. Loomis *et al.* *Lancet Oncol.* <http://doi.org/bj6x>; 2016).

## South Pole CO<sub>2</sub>

Atmospheric concentrations of carbon dioxide at the South Pole have passed 400 parts per million (p.p.m.) for the first time in 4 million years, the US National Oceanic and Atmospheric Administration reported on 15 June. The region crossed this symbolic threshold on 23 May, and was the last place on Earth to do so. Average global CO<sub>2</sub> concentrations, which registered around 280 p.p.m. before the Industrial Revolution, hit 399 p.p.m. in 2015 and are expected to exceed 400 p.p.m. in 2016.

## Earth's pet rock

Astronomers in Hawaii have spotted an asteroid that has been tagging along with Earth for almost a century. The companion, designated 2016 HO3, is estimated to be 40–100 metres in diameter and poses no threat to Earth, NASA said on 15 June. It orbits both our planet and the Sun, drifting ahead of or behind Earth from year to year (**pictured**), but staying within a range of 38–100 times the Earth–Moon distance. Other 'quasi-satellites' have been found, but 2016 HO3 is the most stable



example so far. Quasi-satellites remain close enough to Earth to make them good targets for future spacecraft missions, says Robert Jedicke, an astronomer at the University of Hawaii in Honolulu.

### FUNDING

## Gun-violence crisis

California's state legislature voted on 15 June to establish a US\$5-million firearms-violence research centre at the University of California, following the 12 June mass shooting in Orlando, Florida, that left 49 people dead and 53 injured. There is little research on gun violence; since 1996, it has been difficult — if not impossible — for the US Centers for Disease Control and Prevention (CDC) to study the topic, owing to a federal budget restriction that limits

research that could be used to promote gun control. Although US President Barack Obama ordered the CDC in 2013 to resume the research, the agency says that it does not have the resources. Separately, on 14 June, the American Medical Association declared gun violence a public-health crisis, and resolved to lobby Congress to fund the CDC to investigate the causes of the problem. See page 436 for more.

## Excellence strategy

Germany's federal government and the governments of its 16 states agreed on 16 June to permanently continue a multibillion-euro programme set up in 2005 to strengthen the performance of selected research universities. From 2017, universities will be able to apply for an extra €10 million to €15 million

(US\$11 million to \$17 million) per year from a €533-million annual budget that the governments have earmarked for the extended initiative, renamed the Excellence Strategy. Over the next decade, the windfall will help to create up to 15 'excellence universities', as well as dozens of local and regional research hubs in selected fields of science.

### FACILITIES

## Biggest icebreaker

Russia launched what will be the world's largest and most powerful icebreaker on 16 June in St Petersburg, according to state news agencies. The *Arktika* is currently just a hull, and lacks a superstructure and the two nuclear reactors that will power it, but officials say that it will join the growing fleet of Russian icebreakers by the end of 2017. The vessel is not specifically designed to enable research, but it arrives as Russia seeks to maintain year-round access to the Arctic, and as many nations pursue economic opportunities presented by thinning sea ice in the region.

## ITER retooled

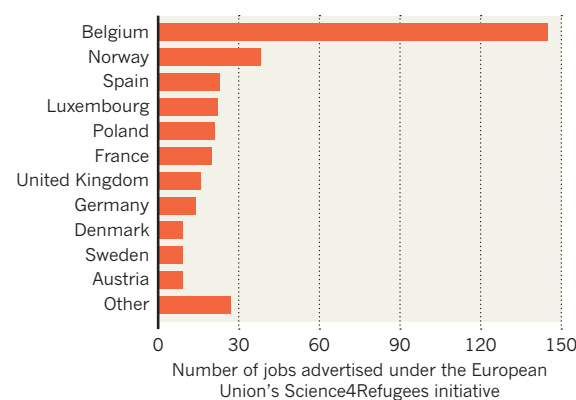
The long-delayed international nuclear-fusion project ITER in St-Paul-lez-Durance, France, is to proceed with a revised construction road map that aims to see 'first plasma' by the end of 2025, its governing board said on 16 June. The streamlined schedule will focus on building a doughnut-shaped vacuum vessel that can generate and confine hydrogen plasma, and postpones the installation of components for sustained fusion of heavy hydrogen isotopes. Sticking to the original schedule would have required an extra €4.6 billion (US\$5.2 billion) from ITER's sponsors. It is not yet clear what savings the new plan will bring in the short term.

## TREND WATCH

Belgian research institutes account for most of the jobs advertised under the European Union's Science4Refugees initiative, which collates posts open to refugee scientists and researchers. The initiative was launched last year to help highly qualified refugees to find jobs. Around 350 adverts from 19 European countries were live on the EURAXESS researcher portal as of 21 June. Of these, 145 were for jobs in Belgium. Refugees compete for the jobs on the same basis as non-refugee applicants.

## REFUGEES WANTED

Belgium offers the most science jobs open to refugees in Europe.



Data as of 21 June 2016.

# NEWS IN FOCUS

**ASTRONOMY** South African community in uproar over telescope **p.444**

**MEDICINE** Infections reveal inequality between the sexes **p.447**

**GRAVITATIONAL WAVES** LIGO spots a second black-hole merger **p.448**



**CHINA** Ten scientists who are shaping the nation's research **p.456**

ROBB KENDRICK/NATIONAL GEOGRAPHIC CREATIVE



The country's lucrative coal industry is one of many things putting climate change on the Australian election agenda.

## POLITICS

# Australian election gives climate researchers hope

*Nation edges towards climate consensus as campaigns avoid carbon controversy.*

BY NICKY PHILLIPS

Australia's elections often feature a fierce debate over climate-change policy. In recent years, arguments over whether and how to put a price on carbon emissions have even swayed voters and toppled governments. Australia is on the front lines of climate change: it is one of the world's largest coal exporters and biggest carbon emitters per capita, and is already experiencing increasingly frequent extreme weather and coral bleaching.

But political uproar over climate change

has been more subdued in the run-up to the election on 2 July, which pits the current Liberal–National coalition government against the opposition Australian Labor Party. In part, that is because politicians are more focused on the country's economy. But policy analysts say the lack of debate suggests that the opposing parties are more closely aligned on climate action this time around. "We're not in as tumultuous a place as we were in previous years," says Frank Jotzo, director of the Centre for Climate Economics and Policy at the Australian National University in Canberra. Both

parties have set emissions targets, and "under both parties something will be done", he says.

A poll of 250,000 Australians, published on 2 June by the broadcaster ABC, suggested that 63% want a price on carbon, up from 50% before the 2013 election. "It is an issue that keeps forcing itself into the conversation," says John Connor, chief executive of the Climate Institute, a policy think tank in Sydney.

At face value, the policies of the two main parties seem distinct. The government, led by Prime Minister Malcolm Turnbull, is promising to continue a scheme that came into effect in

► 2014, which sees companies bid for funding for emissions-reduction programmes. From 1 July, firms will be forced to buy carbon credits if they exceed a ceiling on carbon emissions.

But falling demand for electricity means that is unlikely to happen to electricity generators, says Dylan McConnell, a research fellow at the Melbourne Energy Institute at the University of Melbourne, because the ceiling is set at a high point for emissions that was reached between 2009 and 2014. He says that the government's policies are "definitely not adequate" to achieve its targets of cutting emissions to 26–28% below 2005 levels by 2030. A spokesperson for environment minister Greg Hunt disagreed, saying that the government's policies put it on track to meet its 2030 targets.

The Labor opposition, which under leader Bill Shorten has a slight edge in opinion polls, is more ambitious: it has pledged to reduce emissions by 45% below 2005 levels by 2030, and to be carbon neutral by 2050. Labor would also introduce an emissions-trading scheme for electricity producers. "Labor's policies are stronger with a clearer pathway to credibility than the coalition's, but much detail remains to be sorted," says Connor. For example, it is not clear where the threshold on emissions intensity would be set.

Climate analysts think that in practice, the

two approaches could end up operating in a similar way — with the coalition's pay-to-cut-emissions plan morphing into an emissions-trading scheme similar to the Labor

**"Australia has one of the lowest emissions-reduction targets."**

proposal. "When you look at the policies, unless you're a policy nerd, there's not really much difference," says Tony Wood, head of the energy programme at the Grattan Institute, a think tank in Melbourne. Hunt's spokesperson rebuffed any comparison, saying that the government was not running an emissions-trading scheme.

With warming oceans causing extensive damage to Australia's iconic Great Barrier Reef, rival politicians are keen to be seen as promising action. On 30 May, Shorten pledged Aus\$377 million (US\$279 million) in new funding to improve the health of the reef, if he is elected. Turnbull then announced that his government would use up to Aus\$1 billion from an existing clean-energy programme to support the reef's health through projects to improve water quality, reduce emissions and provide clean energy. But marine biologist Terry Hughes, director of the ARC Centre of Excellence for Coral Reef Studies at James Cook

University in Townsville — who has made headlines with his reef-bleaching studies (see also *Nature* <http://doi.org/bj45>; 2016) — says that the money will make little difference because it won't tackle the fact that global warming is the greatest threat to the reef. "Australia has one of the lowest emissions-reduction targets of any developed country and the highest per capita emissions. Those are the two areas the government should be addressing," says Hughes.

The government also refused to intervene when cuts to climate-change programmes at the national science agency, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), were revealed earlier this year. The opposition has committed to an independent review of the agency and, on 12 June, promised CSIRO Aus\$250 million extra as part of a package to fund various science programmes, if the party is elected.

Jotzo and Wood see the quieter consensus for action on carbon emissions as a relief after a decade of contentious climate politics. Those years saw a carbon 'tax' introduced by Labor's Julia Gillard in 2012, and then dismantled by a conservative coalition led by Tony Abbott in 2014. "There is an opportunity for bipartisanship, which is part of the reason why the toxicity of the debate in this election hasn't been so strong," says Wood. ■

## ASTRONOMY

# Giant telescope rattles rural South African community

*Struggle over Square Kilometre Array highlights balancing act that scientists face.*

BY SARAH WILD, NORTHERN CAPE PROVINCE

"Move it away! We don't want it!" a farmer shouted at a crowded meeting in Carnarvon, a small town in the semi-arid, sparsely populated Northern Cape, one of South Africa's poorer provinces. He was talking about what will be the largest radio telescope in the world, the

international Square Kilometre Array (SKA), a portion of which is due to be built nearby.

Representatives from SKA South Africa, an organization of scientists, engineers and technocrats, were attending the meeting of farmers in May, in an attempt to respond to rising criticism of the project from local people. "It's fine to be part of the international community, but how is it helping this community?"

came a faceless call from the other side of the meeting hall.

In 2012, the SKA's coordinating organization decided that it would divide its thousands of dishes and many more antennas, whose combined 'collecting area' for radio waves will span approximately one square kilometre, between Australia and South Africa. The site in the Northern Cape will include 197 dishes, and



**MORE  
ONLINE**

## TOP NEWS



Man-made pollutants found in Earth's deepest ocean trenches  
[go.nature.com/28lgqkh](http://go.nature.com/28lgqkh)

## MORE NEWS

- The physicist who studies ISIS communities online [go.nature.com/28kujnm](http://go.nature.com/28kujnm)
- Farming invented twice in Middle East [go.nature.com/28o8que](http://go.nature.com/28o8que)
- Cyberattack highlights university vulnerability [go.nature.com/28lhnr](http://go.nature.com/28lhnr)

## NATURE PODCAST



Transmissible cancer; the hadron menagerie; and the latest gravitational-wave result [nature.com/nature/podcast](http://nature.com/nature/podcast)



The MeerKAT telescope under construction in South Africa's Northern Cape will form part of the world's largest radio telescope.

form part of the project's first phase, SKA1. The 64-dish MeerKAT telescope, which will be part of SKA1, is already being built. The rest of the dishes will be added from 2018.

Last year, opposition to the Thirty Meter Telescope on Mauna Kea, Hawaii, prompted the state's supreme court to invalidate the telescope's construction permit. Opposition to the SKA is unlikely to derail the project because legislation protects most of the Northern Cape for astronomy. But SKA South Africa officials say that they need community buy-in if the project is to be sustainable over its 50-year life.

The struggle that is playing out in the Northern Cape illustrates the balancing act that scientists who lead gigantic projects must pull off — to highlight the benefits that the project will bring to an area without over-inflating expectations.

When SKA South Africa proposed the SKA project to the Northern Cape community, starting with the MeerKAT telescope

in 2008, it said that the project would lead to local economic development, create jobs and improve opportunities for children through education and science. But the organization never quantified these objectives — and now its director, Rob Adam, is struggling to manage the expectations of the poorest members of the Northern Cape, who are largely 'coloured' people, a recognized racial classification in South Africa.

SKA South Africa has already come good on some of its promises. It now employs a high-school maths and science teacher for Carnarvon, for example, and is paying for five coloured students at Carnarvon high school to attend university as part of a pan-African bursary programme that it runs. But members of the coloured community complain that such resources haven't materialized across the board — not all the towns in the area have gained a high-school teacher, for example.

And although a small influx of scientists,

engineers and contractors has to some extent improved the economies of the province's towns, the communities are not yet satisfied. "What's in it for us?" asked one resident at a meeting in the Northern Cape town of Brandvlei in May.

Adam says that the community's expectations have risen beyond what the SKA can provide. "You must understand, we are not the government, the education department and the police, all rolled into one," he told the crowd in Brandvlei.

The problem is different for members of the richer, mainly white, sheep-farming community of the Northern Cape, who are concerned about SKA South Africa's land acquisition.

According to the Astronomy Geographic Advantage Act, which was passed in 2007, the government has the right to acquire land for the project within a designated 'core' area if negotiations fail, and if the land is required for the SKA and the organization has offered a fair price.

In 2008, the government bought Losberg farm, the site of the MeerKAT telescope. What is riling this community is that SKA South Africa is now eyeing 36 other farms — which comprise about 118,000 hectares — to accommodate the further 133 dishes that make up SKA1.

Many farmers say that the loss of their farms will destroy the local, agriculture-based economy, and that they are being forced to sell. Although the amount of land needed for the SKA is now agreed, the farmers are also suspicious about the scope of the project. "They don't believe things will stop here," says Henning Myburgh, general manager of farmers' organization Agri Northern Cape in Kimberley.

The spectre of Zimbabwe-style land expropriation, in which the government took land from white farmers without compensation, is also present. "It's a land grab, one way or the



Northern Cape residents voice discontent with construction of the Square Kilometre Array.

► other, be it for SKA or something else,” says Eric Torr, a former resident of the province who owns a local aviation company.

Expropriation would be a last resort, say SKA South Africa officials. “It’s not in the best interests of the SKA to do that because we have to live in this community,” says Alice Pienaar-Marais, who is in charge of the land-acquisition process. She is confident that SKA

South Africa will acquire the 36 farms by the end of next year, in time for SKA1 construction in 2018.

SKA Australia, meanwhile, “could be doing more” with respect to community engagement, project director David Luchetti told *Nature*.

The Australian SKA Pathfinder telescope, which is currently being commissioned, is to be built on an area that traditionally belongs

to the Wajarri Yamatji tribe. Following the 2009 Indigenous Land Use Agreement, which was negotiated between the government and the indigenous group, the tribe has received benefits worth more than Aus\$18.1 million (US\$13.5 million) in exchange for the use of the land for radioastronomy.

But the agreement needs to be renegotiated for the SKA before construction starts. ■



An earthquake-triggered tsunami is thought to have drowned this Pacific Northwest forest 2,000 years ago.

#### SEISMOLOGY

# Canada builds quake warning system

*Undersea instruments will monitor the Cascadia fault zone.*

BY NICOLA JONES

On 15 June, Canada broke ground on its first earthquake early-warning system. Sea-floor sensors will monitor the Cascadia subduction zone off British Columbia to provide crucial seconds of warning if the ‘big one’ hits. Putting sensors so close to the fault should give the Canadian system an edge over a more developed sister project in the United States.

To produce early warnings of quakes,

scientists rely on a network of seismometers and accelerometers to detect the tremor’s first, non-destructive primary (P) waves. Those waves travel faster than the destructive secondary (S) waves, and so hit cities seconds to minutes earlier. The closer that detectors are to the source of an earthquake, the more warning they can provide. That time can be used to stop high-speed trains, shut down nuclear reactors and tell the general population to brace for shaking. But with offshore faults, getting close to the action means putting sensors

under water, which is very expensive.

Japan pioneered earthquake early warnings. The country has had a system to stop bullet trains since the 1960s, and public warnings have been issued since 2007. During the magnitude-9 Tohoku earthquake of March 2011, residents of Sendai, the major city nearest to the epicentre, got 15 seconds of warning; Tokyo got more than a minute. Japan added data from an array of undersea seismometers to its earthquake early-warning system in August 2011, and a second phase of that project was completed in March, more than doubling the number of offshore detectors to 50. Now, the country is working on an ambitious 150-station network called S-net. Connected by 5,700 kilometres of cable, it could provide up to an extra 30 seconds of warning for a large offshore quake.

The United States and Canada have lagged far behind Japan, despite the fact that the Cascadia subduction zone off North America’s west coast is expected to one day produce a catastrophic ‘megathrust’ quake similar to the Tohoku one.

By the end of June, the research non-profit group Ocean Networks Canada (ONC) in Victoria plans to have installed three accelerometers on its NEPTUNE sea-floor observatory, which consists of more than 840 kilometres of ocean-bottom cable looped out past the Cascadia fault (see ‘Quake watch’). “I took this job and asked, ‘Why aren’t we doing earthquake early warning?’” says ONC president Kate Moran, who joined the organization in 2011 as director of NEPTUNE.

The network already has a handful of seismometers, but these send data back in packets instead of instantaneously, and the information is subject to censorship by the navy. As such, Moran says, they are ill-suited for an early-warning system. The new accelerometers, which have a simpler data stream designed to circumvent these issues, were made possible by a Can\$5-million (US\$3.9-million) grant from the British Columbia government in February.

The team is also hoping to install a tiltmeter down a 300-metre borehole, to detect slow, almost imperceptible movement of the tectonic plates at the fault. Clusters of such slow-slip events occurred before the 2011 Japan quake, and detecting them might help seismologists to track the strain that is building on the fault.

Moran anticipates that within 5 years, the ONC will have 40 accelerometers on- and

JOHN STANMEYER/NGC

SOURCE: ONC



offshore to produce public early warnings. The instruments will be positioned specifically to detect an earthquake resulting from a subduction-zone tremor. Spotting quakes from other faults, which would be smaller but potentially much closer to cities, would require a significantly denser network of accelerometers.

On the US west coast, a network of onshore accelerometers can already alert a select group of users — such as the Bay Area Rapid Transit system in northern California — to the early rumbles of earthquakes. That prototype programme, called ShakeAlert, is hoping to get its information to a wider audience soon. “I think we’re really now, finally, at the beginning of rolling out a public system, after years of trying to get funding,” says ShakeAlert lead Richard Allen, a seismologist at the University of California, Berkeley, who anticipates issuing public alerts within five years. ShakeAlert got its first congressional funding in December 2014, and now has about half the funds it needs for a full system, says Allen. To reliably detect quakes from multiple fault lines, Allen reckons that the network needs about 1,100 detectors just in California, where it currently has about 500.

The US National Science Foundation supports a handful of wired sea-floor seismometers off the coast of Oregon as part of its Ocean Observatories Initiative. But these sensors have the same problems as the current Canadian ones, says Martin Heesemann, a marine geoscientist with the ONC. He adds that the group’s new accelerometers will be the only instruments on North America’s megathrust fault designed specifically for early warning rather than research.

The offshore Canadian system “will totally be better” than the US system, says Moran with a laugh. “It’s nice to be better than the United States.” ■

## BIOMEDICAL RESEARCH

# Sexes deal differently with infection

*Quirks of immune system pose medical conundrum.*

BY SARA REARDON

The immune systems of men and women respond very differently to infection — and scientists are taking notice. Research presented last week at a microbiology meeting in Boston, Massachusetts, suggests that the split could influence the design of vaccination programmes and lead to more targeted treatment of illness.

Hints that men and women deal with infection differently have been around for some time. In 1992, the World Health Organization hastily withdrew a new measles vaccine after it was linked to a substantial increase in deaths of infant girls in clinical trials in Senegal and Haiti. It is still not clear why boys were unaffected, but the incident was one of the first such examples to catch scientists’ attention.

Women might have evolved a particularly fast and strong immune response to protect developing fetuses and newborn babies, says Marcus Altfeld, an immunologist at the Heinrich Pette Institute in Hamburg, Germany. But it comes at a cost: the immune system can overreact and attack the body. This might explain why more women than men tend to develop autoimmune diseases such as multiple sclerosis and lupus.

Yet very few studies assess men and women separately, so any sex-specific effects are masked. And many clinical trials include only men, because menstrual cycles and pregnancies can complicate the results. “It’s sort of an inconvenient truth,” says Linde Meygaard, an immunologist at University Medical Center Utrecht in the Netherlands. “People really don’t want to know that what they study in one sex is different from the other.”

Now, scientists are beginning to tease out some precise mechanisms. At the meeting, infectious-disease researcher Katie Flanagan at the University of Tasmania in Australia reported on a tuberculosis vaccine given to Gambian infants. She found that the vaccine suppressed production of an anti-inflammatory protein in girls, but not boys. This boosted the girls’ immune responses, and may have made the vaccine more effective.

Hormones also play a part. Oestrogen can activate the cells involved in antiviral responses, and testosterone suppresses inflammation.

Treating nasal cells with oestrogen-like

compounds before exposing them to the influenza virus has revealed further clues, says Sabra Klein, an endocrinologist at Johns Hopkins University in Baltimore, Maryland. Only the cells from females responded to the hormones and fought off the virus (J. Peretz *et al. Am. J. Physiol.* <http://doi.org/bj5w>; 2016).

Genetic factors may also guide how the sexes deal with infection. Meygaard studies a protein called TLR7, which detects viruses and activates immune cells. Encoded by a gene on the X chromosome, the protein causes a stronger immune response in women than in men (G. Karnam *et al. PLoS Pathogens* <http://doi.org/bj5x>; 2012). Meygaard suspects that this is because it somehow circumvents the process whereby one of the two X chromosomes in women is shut down to avoid overexpression of proteins.

A study set to begin later this year could help to tease apart the relative influence of genes and hormones on infection. Altfeld and his colleagues will look at 40 adults

**“It’s a sort of inconvenient truth.”**

going through sex-change operations. If female hormones are responsible, the transgender women in the study should begin mounting stronger immune reactions to infections and develop more autoimmune problems than the transgender men.

Whether such results will lead to changes in how drugs are administered is an open question. In 2014, the US National Institutes of Health (NIH) announced that researchers must report the sex of animals used in pre-clinical research. Similar efforts are under way in Europe. But a 2015 report from the US Government Accountability Office (GAO) found that the NIH does a poor job of enforcing rules requiring that clinical trials include both sexes (see [go.nature.com/28ll4nb](http://go.nature.com/28ll4nb)).

According to the GAO, even if studies include both sexes, the NIH also does not routinely track whether researchers have actually evaluated any differences between them. Klein argues that gathering such data could lead to more-effective programmes — halving vaccine doses for women, for instance.

“People are tending to ignore it for as long as possible,” Flanagan says. “People will get a lot of surprises.” ■

## EARTH SCIENCE

# Iconic Antarctic lab gets the boot

*Homeless sediment collection dates back to 1960s.*

BY ALEXANDRA WITZE

Free to a good home: more than 23 kilometres of skinny tubes of dirt. With them comes half a century of Antarctic geological history.

The US National Science Foundation (NSF) is looking for a new place to store its Antarctic marine-sediment cores, the world's biggest collection of environmental records from the Southern Ocean. The cores have lain on shelves at Florida State University in Tallahassee since 1963. But last year, the university told the NSF that it no longer wanted to host the collection. Ideas for where the Antarctic Marine Geology Research Facility might move to are due by 3 August.

"This area of research is not a priority for the current faculty," says Gary Ostrander, vice-president for research at Florida State. "It doesn't make sense to continue to support that size facility."

The NSF contributes roughly US\$280,000 per year, but the university has to pay for overhead costs such as air conditioning for the 930-square-metre building.

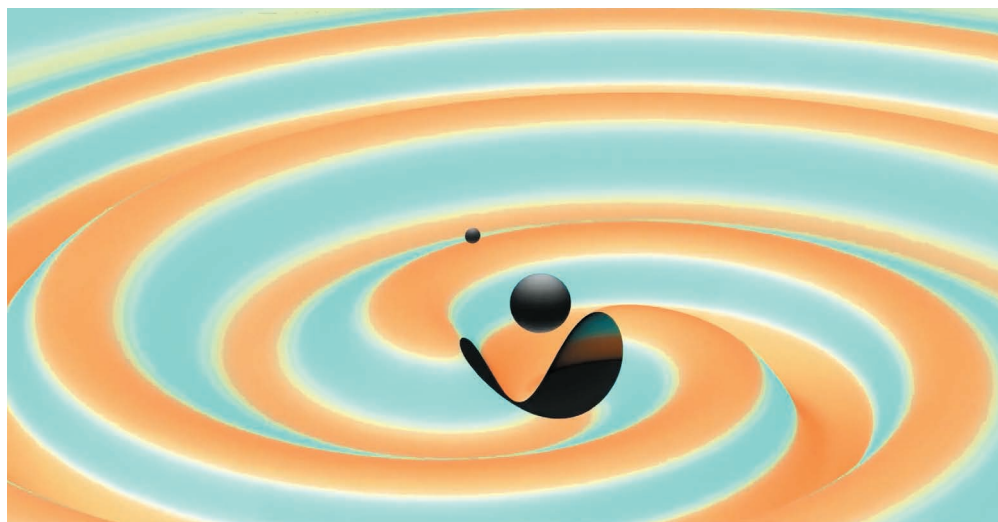
The invaluable collection includes cores gathered in the 2000s by the international ANDRILL programme, which revealed the history of the West Antarctic Ice Sheet over the past 17 million years.

The transfer is a blow for Sherwood Wise, a geologist at Florida State and the facility's principal investigator. "It'll be a sad day for me," he says. "This has been a marvellous resource for the university." Dozens of researchers from around the world visit the collection every year to study palaeoclimate clues and other evidence buried within the sediments.

Over the years, more and more cores have accumulated, from more than 90 research cruises. Studies of the samples have triggered hundreds of publications on all aspects of Southern Ocean and Antarctic history.

Curating these older materials is vital because Antarctic samples are so expensive and difficult to gather, says Philip Bart, a marine geologist at Louisiana State University in Baton Rouge. "The facility is critical to ongoing research," he says.

Wherever and whenever the Florida cores move, Wise estimates that it will take around \$2 million just to pack them up and ship them. ■



A computer simulation of the black-hole merger detected on 26 December 2015.

## PHYSICS

# LIGO sees second black-hole crash

*First gravitational-wave detection was not a fluke.*

BY DAVIDE CASTELVECCHI

Just before 4 a.m. on 26 December, B. S. Sathyaprakash woke up to some good news: gravitational waves had been detected for the second time in history.

The theoretical physicist from Cardiff University, UK, had his laptop next to his bed, set up to alert him when he received automated e-mail notices from computers at the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO).

"I got up and I went and checked the computer. Lo and behold, there was an event from just two minutes before," he says. At 3:38:53 UTC (Coordinated Universal Time), LIGO's twin detectors in Louisiana and Washington state had both picked up the signature ripples of two massive objects — probably two black holes — in the final stages of spiralling into each other.

At the time, the international LIGO collaboration and its colleagues at Virgo, a European observatory near Pisa, Italy, were busy analysing LIGO's first discovery: the event that they had detected on 14 September. The scientists would announce that finding in February, to great global fanfare. They did not run a full analysis of the second event until weeks later, says LIGO physicist Bruce Allen, managing director of the Max Planck Institute for Gravitational Physics in Hanover, Germany.

"It was absolutely mind-boggling how, within a few months of the first event, we had a second one," Sathyaprakash says.


## SECOND SUCCESS

The second detection "shows that this whole business is not a fluke", says Clifford Will, a theoretical physicist at the University of Florida in Gainesville, who is not a member of either the LIGO or Virgo teams. In principle, the September discovery could have been a huge stroke of luck, but the second event suggests that there is a large population of black-hole pairs out there that will produce frequent mergers. LIGO and Virgo can look forward to regular detections, says Will, who studies gravitational waves and other predictions of Albert Einstein's general theory of relativity. "This is going to be a new kind of astronomy."

Einstein predicted that any accelerating or rotating bodies should produce ripples in the fabric of space, which are vaguely similar to sound waves but move at the speed of light and can propagate in a vacuum.

Detailed analysis of the second detection confirmed that the signature had to be the ripples from a pair of black holes (see video at [go.nature.com/28lwdkf](http://go.nature.com/28lwdkf)). This time, the signal from the gravitational waves lasted for one full second, instead of one-fifth of a second as in the first event. The second event encompassed

NUMERICAL-RELATIVISTIC SIMULATION: S. OSSOKINE/A. BUONICCONTI (MAX PLANCK INST. GRAV. PHYS.)/SIMULATING EXTREME-SPACETIME PROJECT; SCIENTIFIC VISUALIZATION: T. DIETRICH/R. HAAS (MAX PLANCK INST. GRAV. PHYS.)



the objects' last 27 orbits around each other, compared with just 5 or so from the first detection. This enabled the researchers to get a test of general relativity that was in some respects twice as precise as their test during the first detection.

This was true even though the September event was 'louder' than the December event. The first two black holes weighed as much as 36 and 29 times the mass of the Sun, respectively, whereas the holes in the second pair were relatively lightweight, at 14 and 8 solar masses, and radiated one-third as much gravitational energy. (In both cases, the black-hole pairs might have been orbiting each other for millions or billions of years, but LIGO captured only the finale, when the orbits and the gravitational waves that they produced had frequencies within the observatory's window of sensitivity.)

### WELL-TIMED GIFT

The LIGO and Virgo scientists estimate that both collisions occurred more than 400 megaparsecs (1.3 billion light years) from Earth, although the distances could not be measured precisely. The teams presented their latest results on 15 June at a meeting of the American Astronomical Society in San Diego, California, and published them in *Physical Review Letters* (B. P. Abbott *et al.* *Phys. Rev. Lett.* **116**, 241103; 2016).

The latest discovery was especially

exciting for physicist Chad Hanna, a LIGO collaborator at Pennsylvania State University in University Park. When Hanna got the alert by text message, it was still Christmas Day in the United States, and he was with family. The collaboration requires members to keep data completely confidential, so Hanna hopped out of his chair, got his laptop and went upstairs to an empty room. At first, he was sceptical, he says: "I just didn't think the Universe had such a sense of humour to send a real event on Christmas."

But he soon realized that although the signal was relatively quiet, it was genuine. It was a crucial test for the software that combs the data from the two observatories in real time, which he helped to design. The system could catch events even if they were half-buried in noise, but didn't produce too many false positives.

### RESET

Ultimately, the automated alerts from LIGO will notify dozens of teams of astronomers as well. The researchers will then reposition their telescopes in the hope of detecting visible light, or other electromagnetic waves, that originate from the same events that produce the gravitational waves.

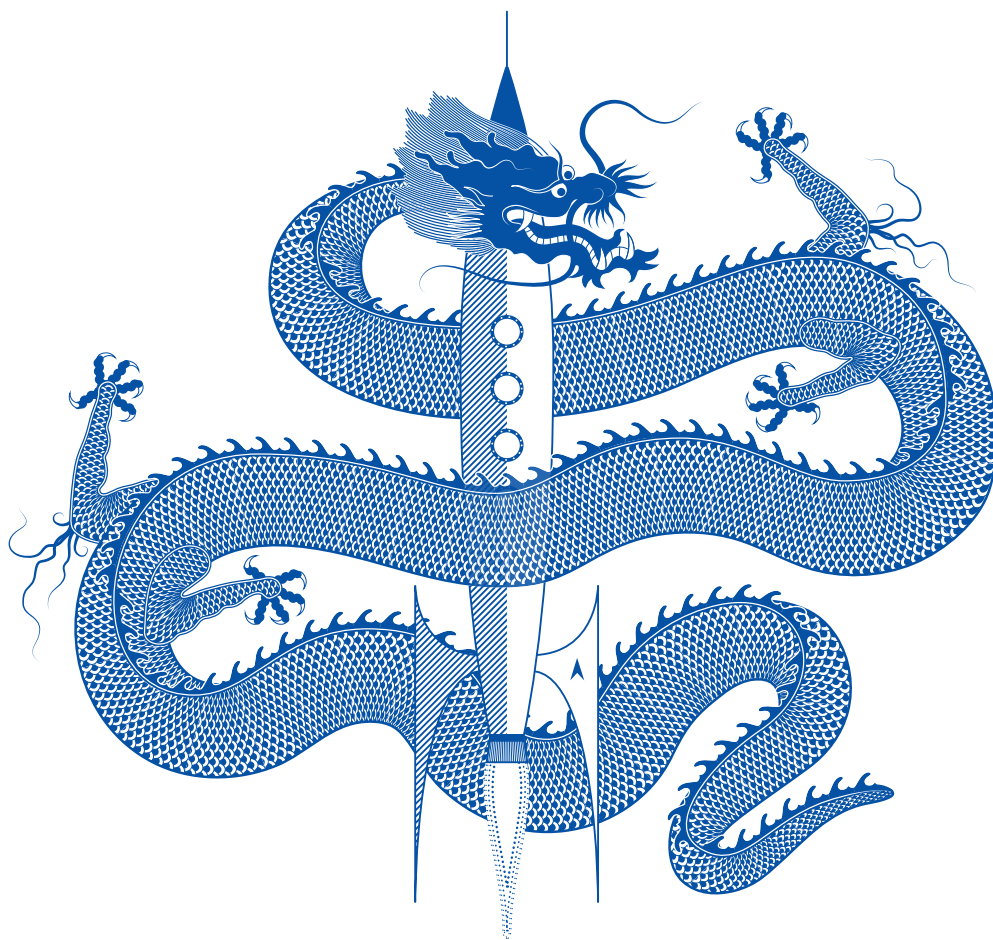
After its first four-month science run between September 2015 and January 2016,

the US\$620-million Advanced LIGO has shut down for an upgrade. It is scheduled to start another run this September, accompanied by an upgraded Virgo.

The results released last week complete the search for black-hole mergers in LIGO's autumn run, but the collaboration is still sifting through its data for other types of event — and may yet announce further discoveries even before the next run begins. In particular, the international Einstein@Home project is looking for signals with the help of computers belonging to volunteers around the world. ■

### CORRECTIONS

It costs the United Kingdom £161 million per week to be in the EU, not £250 million, as stated in the Editorial 'Turning point' (*Nature* **534**, 295; 2016). The News story 'Gene therapies pose million-dollar conundrum' (*Nature* **534**, 305–306; 2016) should have said that cancer drugs that unleash the power of the immune system cost up to \$40,000 per month, not per year. And the News Feature 'The secret history of ancient toilets' (*Nature* **533**, 456–458; 2016) incorrectly said that roundworms and whipworms cause dysentery — they cause problems such as malnutrition.



# Science in China

*A special issue looks at the country's astonishing scientific trajectory as it seeks to secure its spot among the leaders in innovation.*

**C**hinese science has been moving at breakneck speed for the past few decades, fuelled by vast infusions of cash and a rapidly growing technical workforce. China now boasts more researchers than the United States, outspends the European Union in research and development and is on track to best all other nations in its yearly production of scientific papers. But there have been bumps along the way. Chinese research has generally had low impact, and there have been persistent concerns about quality, which the country is trying to address.

This special issue looks at the state of science in China. As part of a new 5-year plan, Chinese leaders have pledged to boost research funding to 2.5% of the country's gross domestic product by 2020. An infographic (page 452) charts the rapid rise of Chinese science and examines some of its problems. On page 456, profiles of ten of the nation's leading scientists show the breadth and

promise of research in fields ranging from neuroscience to neutrinos. One area in which the country is vying to lead the world is DNA sequencing — and an article on page 462 shows that it wants to dominate precision medicine, too.

An Editorial (page 435) notes that even with its impressive scientific gains, China still has far to go before it becomes a leader in innovation. Wei Yang, head of the National Natural Science Foundation of China, which is the leading funder of scientists in the country, says in a Comment (page 467) that China needs to improve the quality, integrity and applicability of its basic research. And policy specialist Douglas Sipp and stem-cell biologist Duanqing Pei argue on page 465 that, contrary to common perceptions, China offers lessons for other nations on how to govern ethically sensitive research in the life sciences. In this and other areas of science, the rest of the world will be watching closely as China races forward. ■

ILLUSTRATION BY JASIEK KRZYSZTOFIAK/NATURE



**SCIENCE IN CHINA**  
A Nature collection  
[nature.com/chinafocus](http://nature.com/chinafocus)

# CHINA

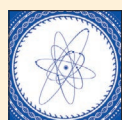
## by the numbers

BY RICHARD VAN NOORDEN

*Research capacity has grown rapidly, and now quality is on the rise.*

China's blazing economic growth has cooled in recent years, but the nation's scientific ambitions show no signs of fading. In 2000, China spent about as much on research and development (R&D) as France; now it invests more in this area than the European Union does, when adjusted for the purchasing power of its currency. That surge in funding has paid off. China now produces more research articles than any other nation, apart from the United States, and its authors feature on around one-fifth of the world's most-cited papers. Top Chinese scientific institutions are breaking into lists of the world's best, and the nation has created some unparalleled facilities.

There's room for improvement within that bright picture. China steers much less of its R&D funding towards basic research than do many science powerhouses, and its international collaboration rates are on the low side. The scholarly impact of its papers is improving rapidly, yet it remains below the world's average. And although China boasts more than 1.5 million researchers, that's a small number given its vast population. The country's leaders recognize some of the weaknesses and have pledged to increase funding for science and technology, aiming particularly to stimulate innovation. ■

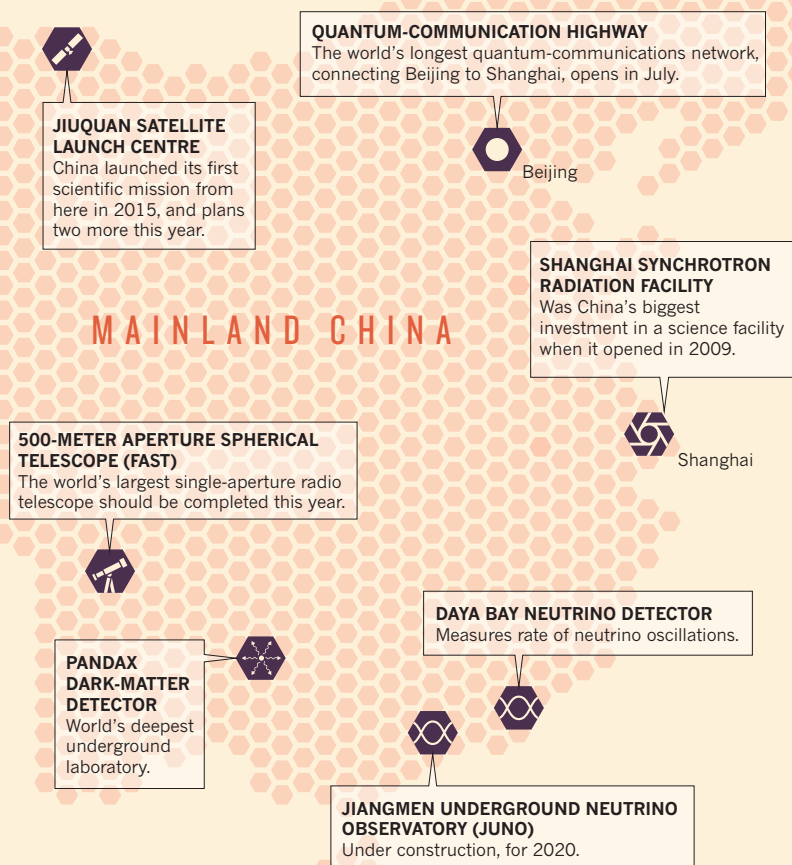


**SCIENCE IN CHINA**

A Nature collection  
[nature.com/chinafocus](http://nature.com/chinafocus)

## BIG SCIENCE

The scale of some of China's experimental facilities — from one of the world's largest telescopes to the deepest underground laboratory — showcases the country's soaring research ambitions.



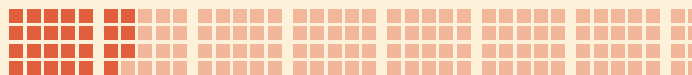
## TOP INSTITUTIONS

The Chinese Academy of Sciences and the nation's leading universities produce tens of thousands of papers every year. A sizeable portion of those rank among the world's best.

■ 250 papers published in 2015 ■ Papers in the world's top 10%

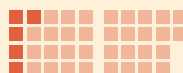
### CHINESE ACADEMY OF SCIENCES

Network of institutes headquartered in Beijing  
36,996 papers / 19% in the top 10%



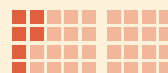
### ZHEJIANG UNIVERSITY

Hangzhou  
9,429 / 14%



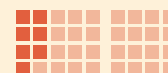
### TSINGHUA UNIVERSITY

Beijing  
9,017 / 18%



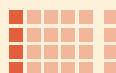
### PEKING UNIVERSITY

Beijing  
9,030 / 20%



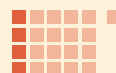
### FUDAN UNIVERSITY

Shanghai  
6,072 / 18%



### NANJING UNIVERSITY

Nanjing  
5,207 / 19%



### UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

Hefei  
4,512 / 22%

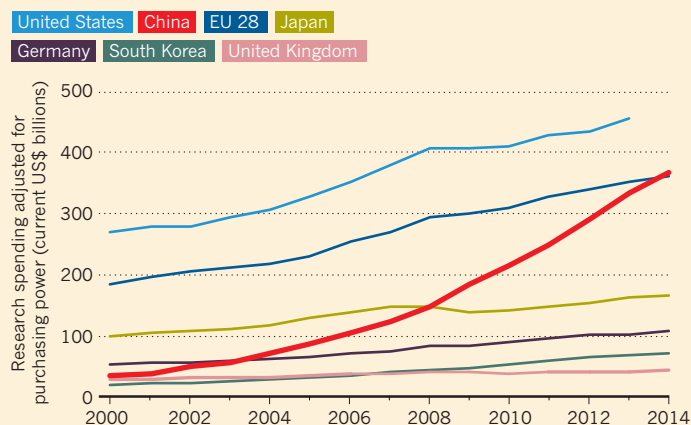


Rounded to nearest 250 papers.

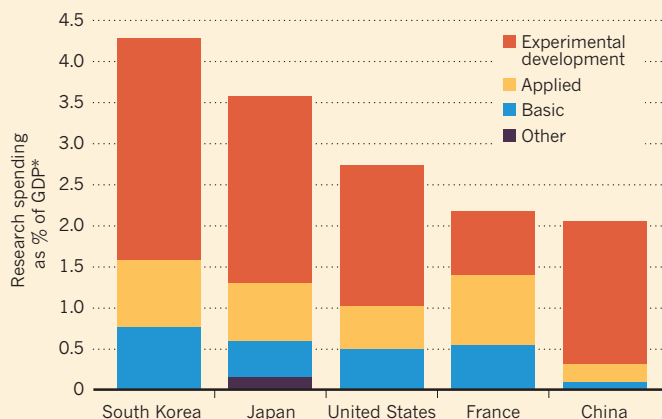
SOURCE: SCIVAL/SCOPUS

## SPENDING

China's science spending is soaring as the country's economy grows and it devotes a greater share to R&D. In absolute terms, China's R&D spending is still only about two-thirds of Europe's. But when its lower wages are taken into account, this translates into a purchasing power that surpasses that of the EU and is on track to overtake the United States by the end of this decade.

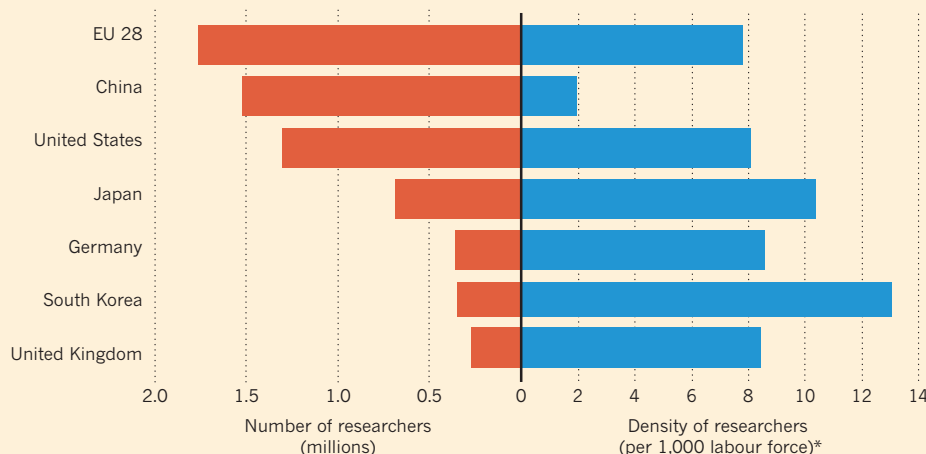


The country now invests more than 2% of its gross domestic product (GDP) on R&D — a greater proportion than the European Union — and wants to reach 2.5% by the end of this decade. However, only 5% of China's R&D spending goes to basic research — a much lower proportion than that of other leading nations. Most of China's R&D funding is aimed at commercially-related technology development.

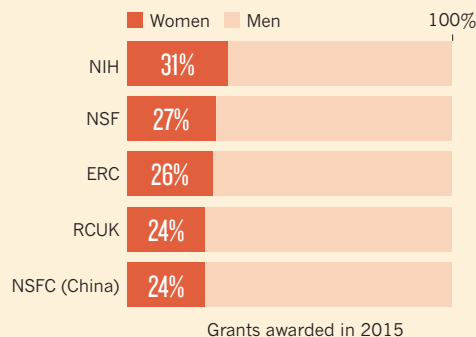


## SCIENCE WORKFORCE

China now has more scientists than the United States. But with a population of more than 1.3 billion, it trails behind other major science nations in terms of the density of its scientific workforce.

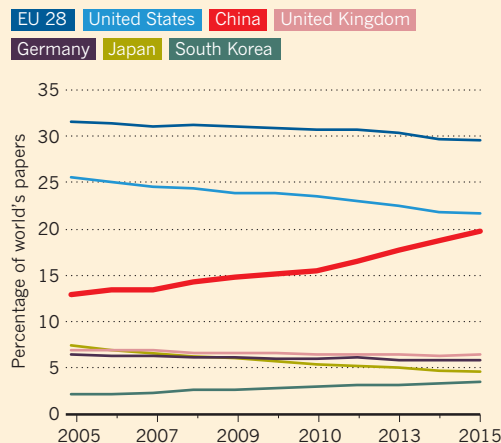


Figures from China's major basic-science funding agency, the National Natural Science Foundation (NSFC), suggest that women now receive around one-quarter of the grants — a figure comparable with that reported by research agencies in the United States and Europe.

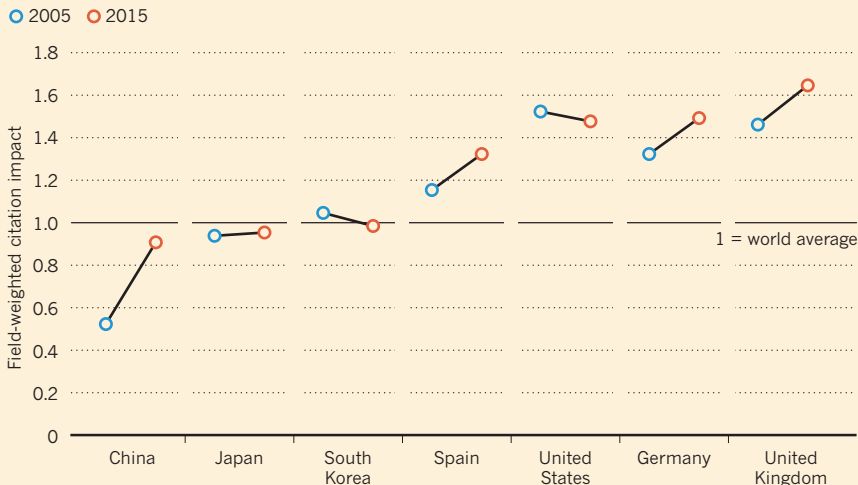


## OUTPUT

In the past decade, China's share of the world's research articles has surged from 13% to 20% — and its share of the world's top-cited articles has shown similar growth.

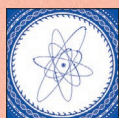


The scholarly impact of the country's output overall remains below the world's average — but it is rapidly improving. The country has its highest impact in the chemical sciences.

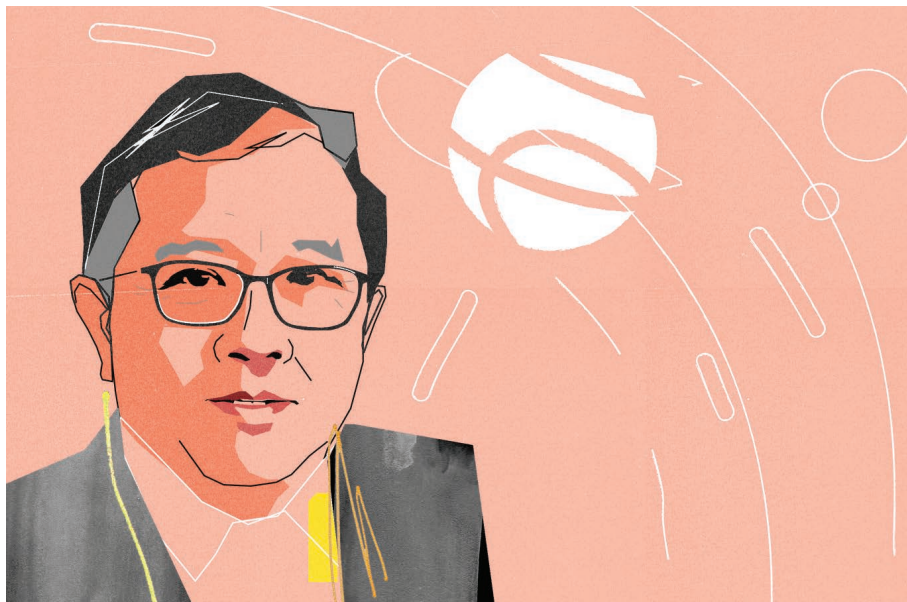


# Science stars of CHINA

From ancient DNA to neutrinos and neuroscience, top researchers in China are making big impacts — and raising their country's standing in global science.



**SCIENCE IN CHINA**  
A Nature collection  
[nature.com/chinafocus](http://nature.com/chinafocus)



## WU JI | UPWARD BOUND

*The country's top space-science official aims high with bold research missions.*

BY CELESTE BIEVER

A fleet of model spacecraft decorates Wu Ji's office, including the Chang'e 3 lander and its Yutu rover that made up China's first mission to explore the Moon's surface. That expedition in December 2013 captivated the world and signalled China's vast ambitions in space. But for Wu, who has been director-general of China's National Space Science Center (NSSC) in Beijing since 2003, a much bigger turning point came almost three years earlier.

On 11 January 2011, he learned that his centre, which is a division of the Chinese Academy of Sciences (CAS), had won funding for a flotilla of spacecraft dedicated to scientific discovery. Up to that point, say Wu and others, almost all of China's space missions had been geared primarily towards advancing national prestige or demonstrating technological prowess.

The 2011 announcement marked the culmination of more than a decade of research, persuasion and international collaboration, mainly on the part of Wu — and the start of a new era in Chinese science. "China has changed direction, and he has been the most important player," says Roger-Maurice Bonnet, former director of science at the European Space Agency, who is an adviser to the NSSC and a scientist at the non-profit International Space Science Institute in Bern.

Two of the NSSC missions have launched. One of them is Wukong, a space telescope hunting for signs of dark matter, which is thought to make up 85% of the matter in the Universe. "The data is coming down every day," says Wu. The mission's team may have an announcement by the end of the year that could "be a mark in science history," he says.

Next up in 2016 will be the world's first space-based experiment to probe the phenomenon of quantum entanglement, and the Hard X-ray Modulation Telescope (HXMT), which will survey a broad region of the sky with greater sensitivity at high energies than other wide-field telescopes.

The funding for these missions has totalled about 3 billion yuan (US\$455 million) since 2011, and Wu succeeded in winning the cash by persuading the top brass at the CAS and China's central government that his agency's proposals for basic space-science missions would deliver breakthroughs. That message resonates with the government's push to invest more in fundamental research.

In person, Wu is hyper-focused on making clear that Chinese research must earn acclaim for its intrinsic value, not just because it is a first for the nation. "There is no Chinese space science," he says. "Only science."

Funding for space research remains a concern because it is allocated in five-year cycles, making it difficult for research communities to mature. But he is confident that space science will gain a steadier source of support — especially if the latest satellites deliver the goods — because both Chinese politicians and the general public increasingly recognize the importance of scientific discovery. "We are a big nation," he says. "For human civilization, we should make contributions." ■

## NANCY IP

**MAKING CONNECTIONS.** Nancy Ip has been building bridges for much of her career. Born in Hong Kong, she found her calling in science during a graduate degree studying neurotransmitters at Harvard Medical School in Boston, Massachusetts. Then she crossed into the biotechnology industry, where she explored the neurotrophic factors that support neuron survival and growth. She took all that expertise back to her native land in 1993, joining the Hong Kong University of Science and Technology (HKUST) when it was just two years old.

"It was considered bold" to move to a place not known for its research, she says, but she wanted to contribute to the region. And since then, she's worked to bolster science and biotechnology there through her research and leadership. "I sleep very little," says Ip, who puts in a 12-hour-plus work day and gives credit to her support team. "Time flies when you are doing things that you enjoy." A lot of that time is spent with her large research group, which spans basic neural biology and translational science for neurological disorders.

Ip has witnessed huge transformations since her return: Hong Kong transferred from British to Chinese control in 1997, and she's seen mainland China's science scene boom. And Ip is now building bridges to the mainland, where she hopes to further clinical research by accessing large populations of people with conditions such as Alzheimer's disease; training people with expertise in both clinical medicine and research; and playing a leading part in a major brain project being developed in China. "I teach my students, sometimes you don't know where research will take you." ■ **BY HELEN PEARSON**





## CUI WEICHENG

**DEEP DIVER.** Cui Weicheng will never forget the dive of his life: riding inside China's *Jiaolong* submersible as it reached a depth of more than 7,000 metres in the Pacific's Mariana Trench 4 years ago. "It's rather desolate down there — but strangely beautiful," says Cui, who led the submersible project.

Thanks to *Jiaolong*, China is now one of only a handful of nations that have the capability to explore the deep sea. *Jiaolong*, which is named after a mythical sea dragon, can travel deeper than any other manned research submersible currently in operation — allowing the country to reach more than 99.8% of the ocean floor.

"This symbolizes China's increasing ambition — and leadership — in deep-sea research," says Jian Lin, a marine geophysicist at the Woods Hole Oceanographic Institution in Massachusetts. Until recently, China's ocean research focused largely on coastal and offshore waters. But, driven by a growing desire for resources and a stronger position in international disputes over marine regions, it is stepping up its support for scientific programmes in the deep ocean.

Now at Shanghai Ocean University, Cui is aiming to reach the deepest place on Earth — the Challenger Deep valley at the bottom of the Mariana Trench, 11,000 metres down. To achieve this goal, he is leading an effort to build a more-pressure-resistant three-person submersible called *Rainbow Fish* at a cost of US\$61 million.

When it is completed in 2020, the vessel will be available for use by scientists around the world, says Cui. "The oceans belong to humanity rather than individual nations." ■

BY JANE QIU

## NIENG YAN CRYSTAL CONNOISSEUR

*A structural biologist unlocks some problem proteins.*

BY ERIKA CHECK HAYDEN

As a girl, Nieng Yan read a classic sixteenth-century Chinese novel featuring a monkey that can transform into other animals. Yan wondered what it would be like to change herself: "If you could shrink yourself into the size of a molecule or a protein, that would be a totally different world," she recalls thinking. Now, as a leading structural biologist, Yan inhabits that world every day, investigating the way proteins work at the level of atoms. "It was almost destined that I would become a structural biologist," she says.

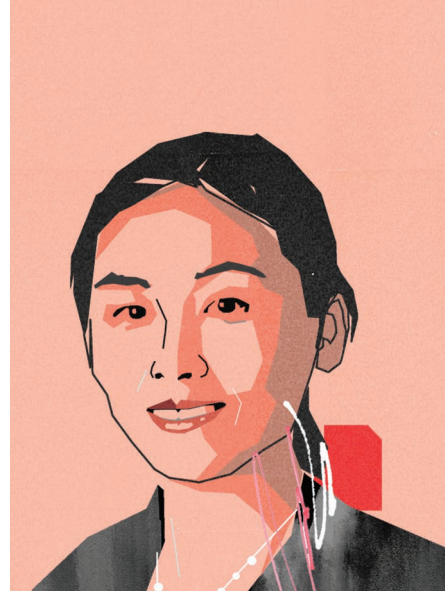
Yan did graduate and postdoctoral research at Princeton University, New Jersey, then set up her own laboratory at Tsinghua University in Beijing in 2007 when she was 30 years old, becoming one of the youngest-ever female professors in China. She focused on determining the structures of proteins embedded in cells' plasma membranes, which are notoriously difficult to solve.

One of her targets was the human glucose transporter GLUT1 — a protein that is essential for supplying energy to cells. Many labs had tried to determine its structure, but the protein had defied their efforts, in part because it readily changes its shape. Yan used a series of tricks to restrict its troublesome movements and finally managed to make crystals and solve its structure in 2014.

"People tried to crystallize GLUT1 for more than 50 years, and all of a sudden, bingo — she hit it," says biochemist Ronald Kaback at the University of California, Los Angeles.

Yan's hits have kept on coming, with a series of high-profile structures. She stays up most nights until 2 or 3 a.m. and skips morning meetings to maximize her time in the lab. Yan has also become a high-profile advocate for better conditions for women and young scientists.

She is excited about using the latest technologies, such as cryo-electron microscopy, which for the first time is allowing researchers to study proteins in fine detail in their native environments, rather than as purified crystals. Yan says that one of the benefits of working in China is she never has to worry about funding and sees a bright future for structural biology there. "The sky's the limit," she says. ■



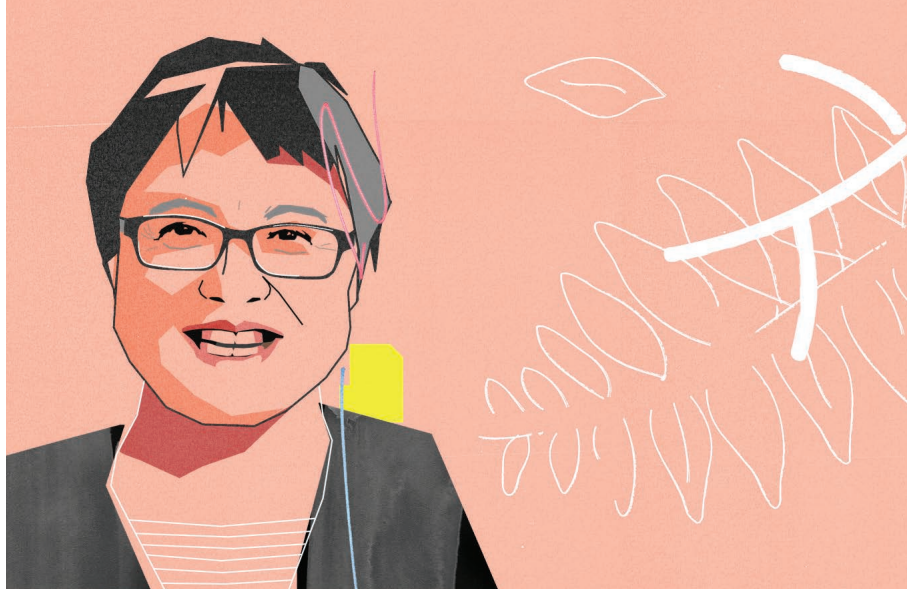
## WANG YIFANG | PARTICLE POWER

*A leading high-energy physicist hopes to smash records with a giant collider.*

BY ELIZABETH GIBNEY

Wang Yifang has a plan to catapult China to the forefront of particle-physics research. The director of the Beijing-based Institute of High Energy Physics (IHEP) wants to build a 50–100-kilometre circular particle collider to succeed the 27-km-circumference Large Hadron Collider (LHC) at CERN, the particle-physics laboratory near Geneva, Switzerland.

The plan is bold, particularly for a country whose biggest existing collider ring is less than 250 metres long. Wang's plan entails building two machines: the first would explore the Higgs



## CAIXIA GAO | CROP ENGINEER

*A gene-editing specialist seeks to make her mark by improving key agricultural plants.*

BY HEIDI LEDFORD

Plant biologist Caixia Gao was initially reluctant to take up gene editing using CRISPR-Cas9 — the technique that is sweeping through biology laboratories around the world. Her lab had already made mutations in 82 genes using an older technology, and the thought of switching to something new was daunting. “At first I felt some resistance,” Gao says. “And then we decided: well anyway, we have to try.”

After a year of frenzied work, her lab at the Chinese Academy of Sciences’ Institute of Genetics and Developmental Biology in Beijing became the first to use the revolutionarily simple gene-editing technique in crops, specifically wheat and rice (Q. Shan *et al. Nature Biotechnol.* **31**, 686–688; 2013). “If there’s any lesson we learn in genome engineering, it’s that you have to be very flexible

and adapt to technology that changes every day,” says Daniel Voytas, a plant biologist at the University of Minnesota in Saint Paul. “Caixia has that ability to adapt.”

She has been doing that for her whole career. Gao went to university planning to go into medicine, but was redirected to agriculture. “Not my interest at all,” she says. “But my thinking is always: as long as I am in this position, I will do my best.” After a PhD in grassland ecology, Gao switched again by taking up plant genetic engineering at the seed company DLF in Roskilde, Denmark.

Gao had to develop methods for inserting foreign genes into grass, which was frustrating work, says Klaus Nielsen, research director at DLF. Many grasses are difficult to engineer, and each species — or even genetic variants within a species — may require its own special

mix of growth conditions. Gao is famously cheerful, but there were days when Nielsen could tell that she was seething.

Even so, she pressed on. “Eventually, she could look in the microscope and see things no one else could see,” Nielsen says. “She was cracking the nut every time.”

During Gao’s 12 years at DLF, she cracked that nut again and again — by genetically altering several traits, including the times when key grass species flower. But European suspicion of genetically engineered crops left her with little hope that her work would leave the lab. “It was so difficult to bring a crop to the market — in the end, the work cannot inspire you any more,” she says. That issue, plus a desire to return with her children to her mother language and culture, sent her back to China.

In Beijing, Gao tackled genetic engineering in wheat, a crop that is legendary for its difficulty to work with, in part because many strains have six copies of the genome. Soon she was considered one of the best in the world at engineering wheat, says Voytas.

Gao is happy with her decision to return to China, where funding for agricultural research is a higher priority than it is in Europe, she says. The government has approved some crops developed with early genetic-engineering techniques, but such approvals have slowed, and China has yet to decide how it will regulate gene-edited crops.

Still, Gao is hopeful that some of her creations will reach the market. Meanwhile, a disease-resistant wheat engineered in her lab is being further developed by a company in the United States. Ever the optimist, Gao refuses to accept public fears about genetically modified organisms (GMOs). “If I meet some people in the street and I ask, they will say they don’t want GMO at all,” she says. “And I stop there and educate them. They are so surprised.” ■

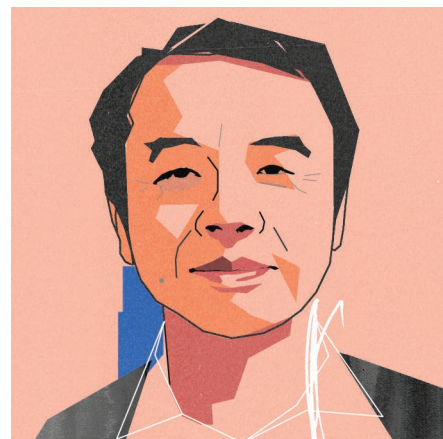
boson starting in around 2028; its follow-up would occupy the same tunnel and smash particles with up to seven times the energy of the LHC.

China will have to compete against CERN, which also wants to host a post-LHC machine. Although China remains the underdog, Wang’s scheme has captured increasing support, says Nima Arkani-Hamed, a theoretical physicist at the Institute for Advanced Study in Princeton, New Jersey, whom Wang brought on board to lead IHEP’s Centre for Future High Energy Physics in 2013. “Now it’s not purely fantasy. It has a chance of really happening,” he says.

Wang says that he only dared to pitch the project because of the success of China’s Daya Bay Reactor Neutrino Experiment. He led that multinational collaboration, which beat international rivals in 2012 by measuring a parameter that governs transformations in the ghostly particles.

At more than 250 times the price of Daya Bay, the Chinese mega-collider will be a harder sell. China’s government has yet to say whether it will foot the facility’s estimated US\$6-billion bill. Brian Foster, a physicist at the University of Oxford, UK, says that Wang has proved he can get major projects off the ground and bring in international support.

And one of his best attributes is persistence, says Shing-Tung Yau, a mathematician at Harvard University in Cambridge, Massachusetts. “He usually succeeds.” ■



# QIAOMEI FU | GENOME HISTORIAN

*A geneticist uses ancient human remains to rewrite Asia's prehistory.*

BY EWEN CALLAWAY

Qiaomei Fu says that she was nervous when she arrived at Germany's Max Planck Institute for Evolutionary Anthropology to pursue a PhD on ancient-human genomics, in 2009. Her master's research in China had focused on the diets of early farmers, and she had no experience with ancient DNA, or even genetics. But Fu jumped headfirst into her new field and "turned out to be one of the most amazing students we've ever had", says Svante Pääbo, a geneticist at the hub for ancient genomics in Leipzig.

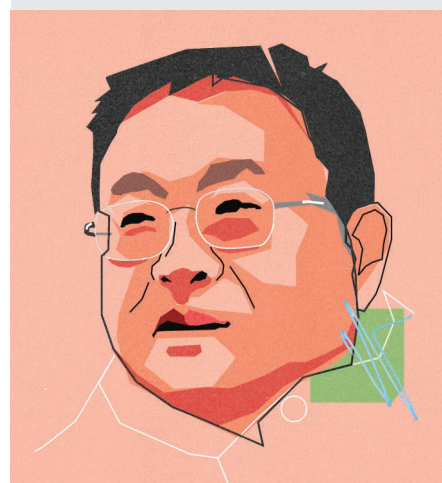
With a trio of *Nature* papers published in the past 20 months, Fu has helped to redraft the history of Europe's earliest modern humans. She returned to China in January to lead an ancient-DNA lab at the Institute of Vertebrate Paleontology and Paleoanthropology (IVPP) in Beijing, where she is set to bring the same upheaval to Asia's ancient past.

She joined Pääbo's team just as it was putting the finishing touches to a draft Neanderthal genome. "It was really high pressure. There were a lot of really interesting things, and a lot of scary things for me," says Fu. "I came there at really the right time." Fu learned how to harvest the scant DNA in ancient bones and quickly picked up evolutionary genetics, bioinformatics and computer programming to analyse the data that she was generating.

Her focus soon turned to the early modern humans who settled Eurasia after leaving Africa, and Fu began collecting and analysing their bones and teeth. She has sequenced the oldest *Homo sapiens* DNA on record: from a 45,000-year-old thigh bone from Siberia and a 40,000-year-old jawbone from a man who had a Neanderthal ancestor in the previous 4–6 generations. Her efforts — culminating in a study of 51 individuals who lived between 14,000 and 37,000 years ago — have shown that Ice Age Europe was more tumultuous than many had thought, with waves of migrants moving in and around the continent and contributing to the ancestry of contemporary Europeans.

Asia's early history may have been even more dramatic than that, because several groups of archaic humans probably coexisted with modern humans, says María Martínón-Torres, a palaeo-anthropologist at University College London who works in China. Fu will turn her attention to the first *Homo sapiens* to settle Asia, who might have arrived more than 100,000 years ago. She also hopes to study Asian history as recent as a few thousand years ago — the IVPP has a vast collection of ancient human bones that have yet to be sampled for DNA.

Fu is often asked why she returned to China instead of staying in the West. "I'm curious what happened in China and east Asia," she responds, "I think it was time to come back." ■



## QIN WEIJIA

**POLAR EXPLORER.** When Qin Weijia first visited Antarctica in 1989, he fell in love with this *terra incognita*. "It's a mysterious continent, full of unknowns and extremes," says Qin, who is the executive deputy director of the Chinese Arctic and Antarctic Administration in Beijing.

Since then, he has been to the frozen continent half a dozen times, including as the 1996 leader of China's first inland traverse towards Dome A — the highest point and one of the least studied regions in Antarctica. That was the first of a series of expeditions, which culminated in the construction of China's Kunlun station on Dome A in 2009.

The country is a relative late comer to polar research, but the Chinese government is investing heavily in both the Arctic and Antarctic, driven by the desire for natural resources and for a bigger say in international discussions about the regions.

Last December, an international team flew ice-penetrating radar and other sensors on China's first fixed-wing aircraft on the continent as it traversed back and forth across thousands of square kilometres over Princess Elizabeth Land in Eastern Antarctica to map features under the ice. "It was the first survey of its kind in a part of Antarctica we know very little about," says Martin Siegert, a glaciologist at Imperial College London. "The results are spectacular."

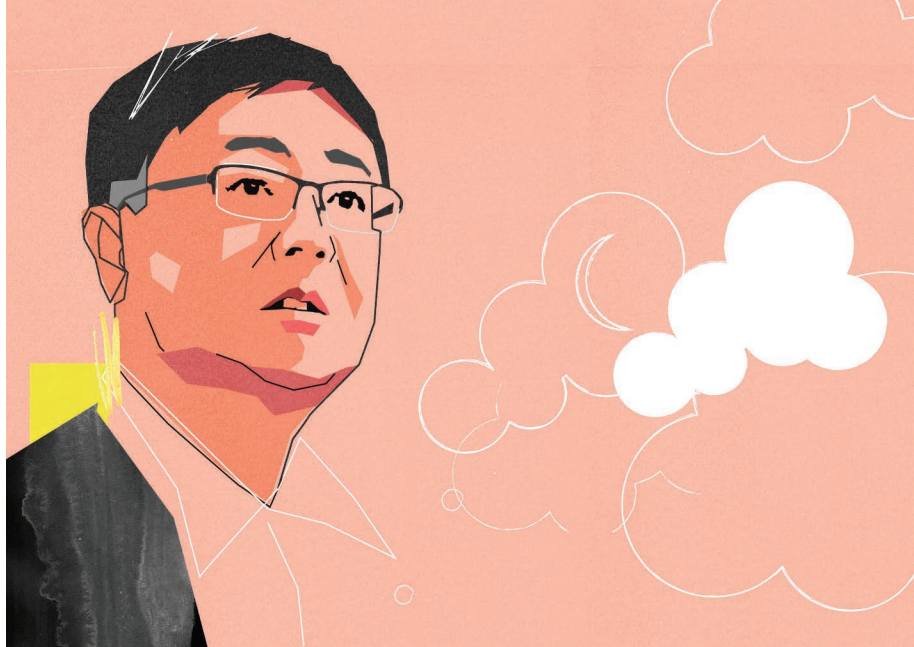
The team discovered the longest canyon on Earth and one of the largest areas of melt under the ice sheet, says Qin, who led the 2015–16 expedition.

Looking forward, he hopes that China will be able to retrieve the oldest ice on the planet from Dome A, which will help to uncover the history of the Antarctic ice sheets and how they have changed. "Only then," says Qin, "can we predict how they will respond to a changing climate". ■ BY JANE QIU

# CHEN JINING

## POLLUTION PATROL

*The top environment official tackles deadly air.*



BY JEFF TOLLEFSON

Chen Jining has a tough job. As head of the Ministry of Environmental Protection, he is responsible for cleaning up the pollution that blankets China's cities, contaminates its drinking water and laces its croplands with toxic compounds. Although he faces formidable odds in one of the most polluted countries in the world, Chen has gained the confidence of many environmentalists and fellow scientists in his first 15 months on the job by stepping up efforts to root out corruption and ensure that local officials and companies are following rules.

"Local officials are being held more strictly accountable on the environment quality," says Li Yan, who is deputy programme director for Greenpeace East Asia and works in Beijing. And because Chen's efforts to reduce air pollution often reduce carbon emissions as well, Li says that the benefits of these reforms extend well beyond the affected areas. "This has massive global implications."

After earning his doctorate at Imperial College London in 1993, Chen worked his

way through the ranks of Chinese academia to become president of Tsinghua University in Beijing in 2012. Now it looks as though he could become the most powerful Chinese environment minister in modern times.

His appointment as head of the environment ministry coincided with a new law that expanded the agency's regulatory powers. He pushed for additional authority to investigate and prosecute polluters, and in May that request was granted. This makes it easier for Chen to intervene when local officials fail to implement many of the government's policies on pollution and development.

In addition to cracking down on pollution, Chen's ministry has worked to strengthen environmental assessments and has boosted transparency by posting more environmental monitoring data on its website, including air-quality readings, as well as information about its enforcement activities.

Chen has often shunned contact with the media, but fellow scientists say that he has been willing to listen to and collaborate with outside scientists and international experts

on issues such as air quality.

"He once said that the history of China's environmental protection is a history of international cooperation on environment and development," says Lailai Li, who heads the Beijing office of the World Resources Institute, on whose board Chen once sat.

The minister still faces huge challenges, however. Citizens are increasingly demanding that the government clean up the environment, but China's rapid industrial rise has created a backlog of problems. Cleaning up the air in major cities may be the easiest task facing Chen; regulators are only beginning to grasp the extent of the water and soil contamination.

And government authorities continue to approve industrial projects, even when the environmental costs are all too clear, says Dasheng Liu, an environmental engineer and research fellow at the Shandong Institute of Environmental Science in Jinan. "He has more power than before," Liu says, but he also faces "more arduous and heavy responsibilities." ■

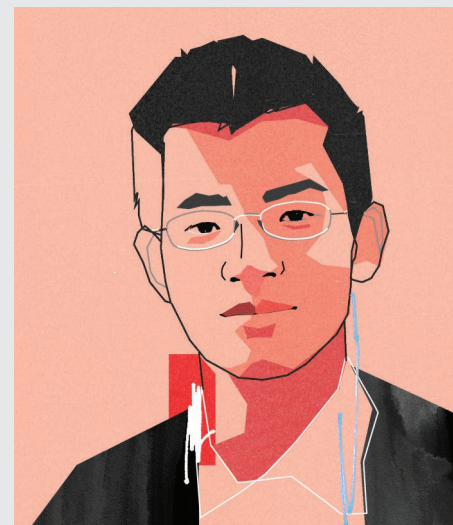
Additional reporting by David Cyranoski.

# CHAOYANG LU

**QUANTUM WIZARD.** When Chaoyang Lu was at school in a tiny village in Zhejiang province, he fell in love with physics. "You could figure out how everything works by a few simple equations," he says.

Now Lu is a rising star in China's push to master quantum information technology — which could eventually lead to powerful new types of computing and secure communications. The 33-year-old, a physicist at the University of Science and Technology of China in Hefei, is noted for his work with 'entanglement', in which the quantum states of different particles are linked regardless of how far apart they are. He has entangled eight photons at once — a world record — and has submitted work using ten. Those achievements led Anton Zeilinger, a quantum physicist at the Vienna Center for Quantum Science and Technology, to call Lu a "wizard of entangled photons". He has also done groundbreaking work with his mentor, Pan Jian-Wei, in the related phenomenon of quantum teleportation, in which a quantum state is transported from one particle to another.

It was Pan who encouraged Lu to do his PhD work at the University of Cambridge, UK, and who convinced him to return to China with the promise that the government is investing heavily in quantum information technologies, and that bright young physicists could focus on research rather than funding. Lu's goal is to advance quantum entanglement enough to use it for computations. "It will be exciting to see, for the first time, a task where a quantum machine can do a better job than a classical one can," says Lu. ■ **BY M. MITCHELL WALDROP**





# THE SEQUENCING SUPERPOWER

First China conquered DNA sequencing.  
Now it wants to dominate precision medicine too.

BY DAVID CYRANOSKI

Six years ago, China became the global leader in DNA sequencing — and it was all down to one company, BGI. The Shenzhen-based firm had just purchased 128 of the world's fastest sequencing machines and was said to have more than half the world's capacity for decoding DNA. It was assembling an army of upstart young bioinformaticians, collaborating with leading researchers worldwide and publishing the sequences of creatures ranging from ancient humans to the giant panda. The firm was quickly gaining a reputation as a brute-force genome factory — more brawn than brains, said some.

Six years later, the scene is quite different. BGI's most famous scientist and visionary leader, Jun Wang, left last July. The machine that had given the company its dominance is outdated, and the firm's attempt to develop its own industrial-scale whole-genome sequencer hit a roadblock last November, forcing it to lay

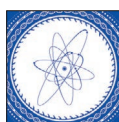
off employees at its US subsidiary. Meanwhile, the competing system — Illumina's X series — has been selling briskly, raising the speed and dropping the price of sequencing worldwide.

Armed with the latest sequencers, rival companies to BGI have emerged. Most prominent of these is Novogene in Beijing, founded in 2011 by former BGI vice-president Ruiqiang Li. And although BGI might not have the uncontested dominance it once did, it still claims to have the world's largest sequencing capacity as well as major scientific ambitions — including to sequence the genomes of one million people, one million plants and animals and one million microbial ecosystems. Today, China is being reborn as a

sequencing power with a broader base.

Fuelling the drive is a multibillion-dollar, 15-year precision-medicine initiative, which China announced in March and which rivals a similar initiative in the United States. If these efforts fulfil their goals, doctors envision being able to use a person's genome and physiology to pick the best treatments for his or her disease. The goal now for sequencing companies is to turn the bounty of genomic data into medical benefits.

To do that, sequence data alone are not enough — so some Chinese companies are going beyond brute-force sequencing to work out how lifestyle factors such as diet are also important for understanding disease risk and for finding therapies. “The thing about China is the ambition they have for their precision-medicine programme is orders of magnitude larger than the United States,” says Hannes Smáráson, chief operating officer and co-founder of



**SCIENCE IN CHINA**  
A Nature collection  
[nature.com/chinafocus](http://nature.com/chinafocus)

SIM CHI YIN/VII

**Scientists at sequencing giant BGI in Shenzhen are looking to apply their genetic expertise to medicine.**

WuXiNextCODE, a genomics company in Cambridge, Massachusetts, that is part of Shanghai-based WuXi AppTec. “They are dynamic and receptive. There, the idea of integrating of genomics into health care is very real.”

## RISE OF THE MACHINE

The new energy behind sequencing is largely thanks to one machine: Illumina's HiSeq X Ten, so called because it is generally sold as sets of ten units. When the machine hit the market in 2014, one set was able to sequence a human genome for close to US\$1,000, and power through some 18,000 human genomes per year. Companies that wanted to rival BGI saw an opportunity — and leapt.

Novogene was the first. Following a model similar to BGI's, Li has been building up a large staff of bioinformaticians to generate and interpret sequence data as part of collaborative basic-research projects on the snub-nosed monkey (*Rhinopithecus roxellana*)<sup>1</sup>, cotton (*Gossypium hirsutum*)<sup>2</sup> and other plants and animals. Using the same machine, a handful of other companies — including WuXi PharmaTech and Cloud Health, both in Shanghai — focus more on offering sequencing as a service to pharmaceutical or personal-genomics companies.

The growth is accelerating. Novogene added a second X Ten set in April, and Cloud Health chief executive Jason Gang Jin says that the company will add another two sets this year. By the end of the year, China will probably have at least 70 units. (Illumina says that 300 units were sold worldwide by the end of last year.)

BGI has been trying to keep pace. In 2013, it purchased Complete Genomics in Mountain View, California, in a bid to create its own advanced sequencing machines for in-house use and for sale. The firm announced a system called Revolocity, its attempt to match the HiSeq X, last June. But in November, having taken just three orders, it suddenly suspended sales. BGI is now left with its ageing fleet of 128 Illumina HiSeq 2000 machines and a mélange of newer sequencers from various companies, including its own.

Estimates of China's share of the world's sequencing-capacity range from 20% to 30% — still lower than when BGI was in its heyday, but expected to increase fast. “Sequencing capacity is rising rapidly everywhere, but it's rising more rapidly in China than anywhere else,” says Richard Daly, chief executive of DNAnexus in Mountain View, which supplies cloud platforms for large-scale genomics.

BGI has another machine up its sleeve. The BGISEQ-500 is designed as more of a desktop instrument for research labs. It is also based on the Complete Genomics technology and is set to begin shipping this year. Yiwu He, BGI's new global head of research, says that the system

can sequence a human genome for \$1,000, and by being smaller in scale and more flexible to use, it will meet China's emerging need for clinical sequencing. “There will be more sequencing done outside of research institutes, in the hospitals,” says He. The company will bring the price of one human genome sequence down to \$200 in the next few years, he predicts boldly. “China is the most exciting place to do biomedical research.”

## GENOMES EN MASSE

The announcement of the precision-medicine programme sent a ripple of excitement through China's sequencing giants. The money will be spent on improving technologies, sequencing, and sharing and analysing more than one million human genomes, as well as on developing drugs and diagnostics from the data and using those findings to personalize medical care. Hungry for a share of the cash, hospitals and clinicians are teaming up with sequencing companies to come up with proposals for the work.

The one million human genomes will be split among a variety of studies, and will include groups of 50,000 people who each have metabolic disease, breast cancer, gut cancer or another condition. There will also be cohorts

**“SEQUENCING CAPACITY IS RISING RAPIDLY EVERYWHERE, BUT IT'S RISING MORE RAPIDLY IN CHINA THAN ANYWHERE ELSE.”**

that represent northern, central and southern China, “to look at the different genetic backgrounds of subpopulations”, says Li.

Similar projects exist elsewhere, including one in the United Kingdom that is sequencing 100,000 genomes, and one in the United States that has a budget of \$215 million and aims to cover one million genomes. But China will have some advantages, observers say, not least of which is firm backing from the government. Over the next five years, the government has promised to add several precision drugs and molecular-diagnosis products to the national medical-insurance list, ensuring that companies' research costs will be recouped if they lead to such a product. In the United States, biotech companies with new products can struggle to get insurance companies or the government to pay. “There is greater acceptance of sequencing and willingness to invest in it in China,” says Daly.

In September, BGI will open the China National Genebank, a five-hectare facility in Shenzhen that will house millions of samples from people, animals, plants and microbes. Entrusted to BGI by the central government, the bank will make some samples and data available to researchers around the world. And the company is compiling its own database of one million human genomes, which will overlap to some degree with the national project. BGI will hit the target before anyone

else in the world, predicts He. “We can get there faster because of our partnership with the government, hospitals, universities, because we can move faster than large consortia, and particularly because we have our own sequencer. That is a huge advantage,” he says.

But making sense of one million human genomes is a major challenge, says Wang, who quit BGI to found a company called iCarbonX in Shenzhen. The firm plans to collect sequencing data for more than a million people “as a start”, as well as other biological information, including changes in levels of proteins and metabolites, brain imaging, biosensors to monitor things such as glucose levels and even the use of smart toilets that will allow real-time monitoring of urine and faeces. Wang calls it a “digital form of you”. He plans to use artificial intelligence to integrate all the data, with the ultimate aim of providing medical care that is tuned to an individual's genes and physiological state. Less than a year in, Wang has raised more than \$100 million, including a big chunk from Shenzhen-based Tencent, the company behind the social-media application WeChat, which Wang says will help to build the data-collection platform.

China is already exploring how else genomics can benefit health. In March, BGI celebrated its one millionth NIFTY test — a screen that sequences fetal DNA circulating in the mother's blood to detect chromosomal abnormalities such as Down's syndrome<sup>3</sup>. The country's conversion from a one-child to two-child policy is expected to accelerate the demand for such tests. Cancer genetics is also well on its way. Cloud Health last year fed genomic data from some 15,000 tumour samples to more than 100 genetics companies in China to help with diagnosis and make sure that patients get the right chemotherapy drugs. The market for pricey genomic tests is growing in step with the country's middle class. “China has 100 million people making more than \$50,000 now,” says Daly.

For Wang, sequencing on its own is old hat. “Genomics is important, but it's just one piece of the puzzle,” he says. “All the complex traits. All the neurodegenerative disorders, cancer, diabetes — it's all more than genetics. If we only talk about genomics, about massive data without clinical info, that's not enough.” ■

**David Cyranoski** writes for Nature from Shanghai, China.

1. Zhou, X. *et al.* *Nature Genet.* **46**, 1303–1310 (2014).
2. Zhang, T. *et al.* *Nature Biotechnol.* **33**, 531–537 (2015).
3. Dan, S. *et al.* *Prenat. Diagn.* **32**, 1225–1232 (2012).

# COMMENT

**CHINA** To-do list from president of National Natural Science Foundation **p.467**

**ECONOMICS** Will one metric or a dashboard topple GDP? **p.472**



**CONSERVATION** Objections to banking African rhinos in Australia **p.473**

**EQUALITY** Heads of research agencies commit to support women **p.473**

REUTERS



Baby boom: following the end of China's one-child policy, the private fertility sector needs close scrutiny.

## No wild east

China has lessons for the world when it comes to overseeing ethically sensitive research in the life sciences, argue **Douglas Sipp** and **Duanqing Pei**.

The first and only published papers to describe genome modification in human embryos have come from Chinese laboratories<sup>1,2</sup>. For some, this is another signal of China's successful transformation from a closed society focused on farming and the manufacturing of commodities to a world leader in innovation. For others, these studies are the latest in a list of feats reported over the past decade that reflect the country's lax regulation or cultural indifference to fundamental ethical tensions.

In our view, fears that China's scientific ambitions are overwhelming its ability to exercise appropriate caution in the life sciences — particularly in research involving human embryos — are overblown. In fact, China has shown care and restraint with respect to altering the genomes of human eggs, sperm or embryos, and in the use of human embryos in research more broadly.

Major challenges lie ahead, particularly in the commercial application of biotechnology. But as international standards evolve to keep pace with rapid advances in research, China should be encouraged to take its place as a fellow pioneer alongside longer-established research superpowers — both in the laboratory and in regulation.

### TRIAL BY MEDIA

The first study to report the modification of the genomes of human embryos was rejected by *Nature* and by *Science*, reportedly in part because of peer reviewers' ethical apprehensions. And media accounts of both papers in the United States and Europe often depicted the work as the pursuit of progress unchecked by principle (see [go.nature.com/1ukutpw](http://go.nature.com/1ukutpw) and [go.nature.com/1tmicqx](http://go.nature.com/1tmicqx)).

It is not unusual for research led by labs in China to be cast in such a light. In the early 2000s, Chinese investigators transferred<sup>3</sup> the nuclei of human skin cells to cultured rabbit egg cells in an attempt to produce humanized stem cells. Like the first gene-editing paper, that study was initially rejected partly because of ethical



**SCIENCE IN CHINA**  
A *Nature* collection  
[nature.com/chinafocus](http://nature.com/chinafocus)

► concerns and received intense media scrutiny when published. More recently, BGI, a private genomics company in Shenzhen, sought to associate particular genetic sequences with intelligence by sequencing the genomes of volunteers with ‘high cognitive ability’. The study, which began in 2012, provoked concern about a supposed interest in eugenics in China despite assurances from the institute that the aim was purely to further basic understanding of the genetic basis of high IQ and that, in any case, reproductive applications would not be allowed under existing guidelines.

**“China’s approach has arguably been more effective and enabling than the legal patchwork seen in much of the world.”**

Studies such as these would rightly provoke discussion wherever they were conducted. But all too often the intimidation is that Chinese scientists are free to do anything and are a step away from making designer babies. What is more, commentators, both in China and outside it, often assume that scientists and others in China have little concern about the fate of early human embryos.

Even a cursory review of China’s existing regulations, as well as its research and social norms, shows that this picture is fundamentally inaccurate.

## PRINCIPLED PROGRESS

National guidelines on assisted reproduction and embryonic-stem-cell research<sup>4,5</sup> have precluded the implantation of modified human embryos for reproductive purposes since 2001. In China, going against government guidelines can incur financial penalties and loss of employment as well as loss of funding and licences to do research. Thus, although not encoded in law, the ruling that research on human embryos is permitted, but that the transfer of modified embryos to a woman’s uterus is not, has been described as a “Rubicon” for China’s research and medical communities<sup>6</sup>. It has not been breached in the 15 years since the 2001 guidelines were written.

Both of the labs that described genome editing in human embryos<sup>1,2</sup> obtained approval for their studies from institutional review boards. They also used non-viable one-cell embryos that had been discarded by *in vitro* fertilization (IVF) clinics and that were incapable of developing to term. Furthermore, they discontinued their work on discovering that the genome-editing process was unexpectedly inefficient. None of these steps suggests a cavalier approach to research involving human embryos.

There is also little evidence for a nonchalant attitude among Chinese citizens towards the use of human embryos in research.

Contrary to common perceptions, families using IVF in China have generally been conservative in their handling of surplus embryos. Even under the one-child policy (a public policy that was lifted only in January this year), 83% of surveyed Chinese families using IVF opted to keep surplus embryos in storage for 0–3 years after the birth of their baby<sup>7</sup>, with many citing feelings of attachment to the embryos as their reason (see ‘To store or not to store’). Around 63% of families similarly surveyed in the United States chose to keep embryos in storage for 0–5 years after having a baby using IVF<sup>8</sup>.

Getting a handle on China’s complex regulatory systems can be daunting for non-Chinese speakers and, when it comes to implementation, sometimes even for the domestic community. Indeed, China would almost certainly earn more international regard if it made more effort to publicize its regulatory framework. Many of the nation’s guidelines are buried in hard-to-navigate agency websites, and official English translations are scarce, making informed discussion by foreign scholars difficult. Yet, in relation to the use of human embryos in research, China’s approach has arguably been more effective and enabling than the legal patchwork seen in much of the world.

For years, stem-cell researchers in the United States have faced uncertainty over the future of the field and over whether a given cell line would remain usable in federally funded work. That confusion has been compounded by differences between

states; currently, the use of embryonic stem cells may be legal in one state but a crime just across the state line. Other countries, including Australia, Brazil and Japan, have similarly struggled to develop embryo-research policies<sup>9</sup>, leaving scientists in limbo for many years.

The binding force of government guidelines, combined with China’s consistent position — prohibiting uses in reproduction but permitting those in research — has given scientists the confidence to pursue studies in a well-defined ‘safe space’. Indeed the significant gains that China has made, particularly in relation to gene editing<sup>1,2,10</sup>, are thanks in part to this clarity.

## PRIVATE SECTOR

A major question now is whether ministerial guidelines will suffice in China’s nascent private biotech sector.

In 2009, 2012 and 2015, the Ministry of Health (MOH) and the China Food and Drug Administration (formerly the SFDA) introduced guidelines for stem-cell therapies. Early attempts to rein in private clinics ran into difficulty, in part due to the diversity of regulatory jurisdictions across regions and cities, and inconsistencies in compliance. Wherever there are ambiguities, enforcement can become challenging. Stem-cell biology is a young science and its clinical application is a therapeutic frontier, and so both regulators and the regulated lack experience. Indeed, the process of ensuring that guidelines are being followed has often become a matter of discretion for individual government agencies — a situation that is ripe for abuse.

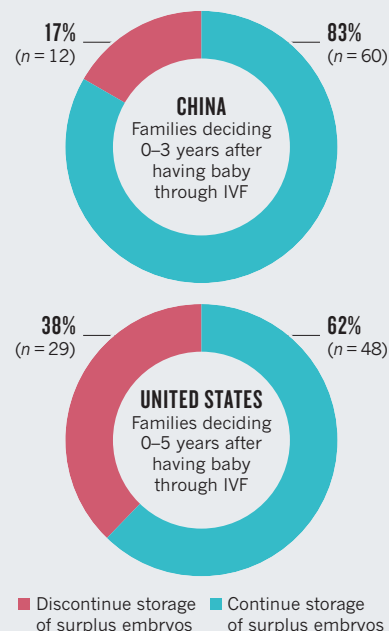
The National Health and Family Planning Commission and China Food and Drug Administration this year established a panel of experts to evaluate centres that seek to perform stem-cell-based clinical trials. Guidelines published by these bodies state that any organization wanting to pursue such trials should first undergo this evaluation (see [go.nature.com/1tvjcw](http://go.nature.com/1tvjcw); in Chinese).

The death in April of a cancer patient who had received an ineffective cell therapy caused an outcry on social media in China, which prompted a government order to hospitals not to outsource medical services. The new evaluation procedures, combined with an increasingly aware public acting as watchdog, should make it harder for providers to market unapproved treatments.

Another concern is that rising demand for reproductive medicine following the relaxation of the one-child policy may expand markets for IVF and other services (see [go.nature.com/1uh4rep](http://go.nature.com/1uh4rep)). Currently, all IVF clinics require a licence in China. But a potentially more profitable private sector in this area warrants close scrutiny. New laws

## TO STORE OR NOT TO STORE

Families in China and the United States who have had babies using *in vitro* fertilization (IVF) make similar choices about whether to continue to store frozen surplus embryos.



SOURCES: REFS 7, 8

EPA or guidelines may be needed to check the marketing of certain services.

Many countries face the challenge of developing effective policies that both respect the ethical standpoint of diverse publics and enable the exploration and application of biomedical technologies. China should be given an equal voice in the global discussion about how best to achieve this. Encouragingly, more dialogue is starting to happen. In an unprecedented move, the Chinese Academy of Sciences joined the US and UK scientific academies in organizing the first international summit on human gene editing last year.

Establishing appropriate governance for research in the life sciences is hard for everyone given globalization, the pace of technological advances, the complexity of domestic regulatory ecosystems and a growing international movement to make deregulated markets — not government officials or bioethicists — the arbiters of quality and ethicality. We must therefore strive for a better understanding on all sides of the efforts that different countries are making, and of how they can work together to develop a consensus on international governance. Good rules drive good science. ■

**Douglas Sipp** is a researcher at the RIKEN Center for Developmental Biology, Kobe, Japan, and project professor at the Keio University School of Medicine, Tokyo, Japan. **Duanqing Pei** is professor at the Key Laboratory of Regenerative Biology, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health (GIBH), Chinese Academy of Sciences, Guangzhou, China, and is director-general of the GIBH.

e-mails: [sipp@cdb.riken.jp](mailto:sipp@cdb.riken.jp);  
[pei\\_duanqing@gibh.ac.cn](mailto:pei_duanqing@gibh.ac.cn)

1. Liang, P. *et al.* *Protein Cell* **6**, 363–372 (2015).
2. Kang, X. *et al.* *J. Assist. Reprod. Genet.* **33**, 581–588 (2016).
3. Chen, Y. *et al.* *Cell Res.* **13**, 251–263 (2003).
4. *Guidelines on Human Assisted Reproductive Technologies* (People's Republic of China Ministry of Health, 2001); available at <http://go.nature.com/1ztc8qb> (in Chinese).
5. *Guidelines on Human Embryonic Stem Cell Research* (People's Republic of China Ministry of Health, 2003); available at <http://go.nature.com/28joycl> (in Chinese).
6. Doring, O. *Kennedy Inst. Ethics J.* **14**, 39–46 (2004).
7. Jin, X. *et al.* *BioMed Res. Int.* <http://dx.doi.org/10.1155/2013/934567> (2013).
8. Nachtigall, R. D. *et al.* *Fertil. Steril.* **92**, 2094–2096 (2009).
9. Kawakami, M., Sipp, D. & Kato, K. *Cell Stem Cell* **6**, 415–418 (2010).
10. Gao, F., Shen, X. Z., Jiang, F., Wu, Y. & Han, C. *Nature Biotechnol.* <http://dx.doi.org/10.1038/nbt.3547> (2016).



The Five-hundred-meter Aperture Spherical Telescope in Guizhou is due to be completed in September.

# Boost basic research in China

Improving the quality, integrity and applicability of scientific research will underpin long-term economic growth, writes **Wei Yang**.

China's economy relies on innovation. Developing technologies, improving efficiency and creating and implementing new scientific knowledge can invigorate industry and help society. China's recent economic slowdown, however, calls for a gear change in how the nation innovates.

For several decades, short-term and focused technological research and development (R&D) has been the main driver in China. Large public grants were channelled to promising or urgent areas to deliver new turbine engines, high-speed trains, solar panels or drugs in 5–10 years. Now China must take a longer and broader view, and nurture its science roots.

Basic research — studies that create scientific knowledge and technologies that can be subsequently developed, translated or applied — has a conflicted image in China. Progress has been enormous (see *Nature* **481**, 420; 2012): China's share of research papers worldwide (as counted in Elsevier's

Scopus database) grew from 2.5% in 1997 to 18.8% in 2015 — but severe criticisms persist (see ref. 1 and *Nature* **463**, 142–143; 2010). For example, critics say that China's universities have become paper mills induced by metrics that value quantity over quality. Impact remains low: few chemical reactions or processes are named after Chinese scholars, even though the nation now publishes more papers in chemistry than any other. Research misconduct — including ghostwriting and reviewing — has been rife, as evidenced by retractions of papers by Chinese authors from BioMed Central, Elsevier and Springer journals in the past two years.

Industrialists and some government officers complain that many academic studies, such as in pure mathematics or fundamental physics, are irrelevant to the nation's economy or society. Scientific and technological progress contributed to only 55% of economic growth in China in 2015, compared with 88% in the United States in the same period. And China spends relatively little of its total R&D budget (public, industrial and private) on basic research — just 4.7% in China compared with 24.1% in France, 17.6% in the United States and 12.6% in Japan in 2013.

Improving the quality and integrity ▶



**SCIENCE IN CHINA**  
A Nature collection  
[nature.com/chinafocus](http://nature.com/chinafocus)

► of basic research must be the focus of national efforts to boost innovation in China. Quality needs to matter more than quantity, and integrity is the best way to ensure quality. Applicability to technological development justifies drawing more resources into basic research. As the head of the National Natural Science Foundation of China (NSFC), a leading national government funding agency for basic-science research, I call for a sustained focus to bring about such a change by 2020.

### EMERGING POWERHOUSE

China is rising rapidly up the global scientific ranks by every measure — quantity and quality of output, R&D spend and increased collaboration (see page 452). For example, China's share of high-impact works (the top 0.1% of papers in Scopus rated by citations) has grown, from less than 1% in 1997 to about 20% now.

**"Raising the bar on quality must be the top priority."**

Founded in 1986 with a starting budget of just 80 million yuan (US\$12.2 million), the NSFC has expanded more than 300-fold to allocate 24.8 billion yuan in 2016. It funded 62.1% of Chinese research papers, or 11.5% of global academic output, in 2015. The foundation's mission is to be a 'FRIEND' of scientists: fair in reviews; rewarding in fostering research; international in global participation; efficient in management; numerous in grants; and diversified in disciplinary coverage.

But beyond the buoyant statistics, basic research in China has been slow to develop. For example, there is only one science Nobel laureate from mainland China. And the nation's research lags behind other countries in terms of citations — its Field Weighted Citation Impact measure was 0.86 in 2015, below the world average of 1.0.

Raising the bar on quality — higher citations and more major breakthroughs — must be the top priority. Put another way, China needs to raise the altitude of its basic research landscape and form high mountains.

Agreed metrics are needed to track progress. Current measures are heterogeneous and do not work equally well across China's vast and diverse academic landscape. The country has 1,000 research institutions capable of basic research, each with a different focus, and more than 1,000 universities, each with a different blend of research and teaching. For example, Tsinghua University in Beijing receives nearly 5 billion yuan in annual research grants from all sources, whereas some regional colleges have research budgets of only a few million yuan per year. Measuring publication numbers might work well for a young institute that publishes ten papers a year in relevant international journals, but may eventually distort the

disciplinary mix of a large university that publishes 10,000 papers a year in diverse journals.

So, in practice, each institute must decide for itself what is most important to track. It might choose to look at whether a project or person is producing many publications, whether the work has high impact or is highly cited, if a project is globally significant or is a major scientific breakthrough. Each institute must plot a trajectory that is consistent with its history, current status and future goal. Evaluation needs must be reassessed as a project matures, and as the institute upgrades. In most cases, when institutes are managing their progress healthily, this 'soft' approach will work. Interventions such as campaigns to reward high-quality work might be needed for those that deviate from the research commonwealth.

Universities need to implement metrics wisely and clarify their aims (see D. Hicks *et al. Nature* **520**, 429–431; 2015). The Chinese Academy of Sciences (CAS) has taken a lead: some 15 years ago, it was the first in China to include citation in its assessment metrics, leading to an exponential growth in high-impact works. And three years ago, CAS directed each of its 104 research institutes to concentrate on one mission, three near-future breakthroughs and five long-term directions.

But setting targets that are too rigid can skew or hinder research. More institutes recognize that emphasizing publication numbers pressures researchers to write lots of incremental papers rather than a few good ones. Merit-based academic evaluations — that account for international recognition, representative works and impacts to the field — can avoid this. Long-term development, which may be slow but steady, must be distinguished from short-term gains that lack sustainability. Many universities and research institutes are downsizing the proportion of researchers' salaries that are based on performance (from more than 70% to less than 30%

in extreme cases), so that a higher percentage goes towards rewarding merit.

Another question is how best to apportion basic research funds. Should science address societal 'grand challenges' or test bold concepts? Should resources be pooled or shared among many individuals? One answer is to cover several bases. For example, the NSFC invests 70% of its funding in blue-skies research, 10% on supporting talent and 20% on major research projects for scientific challenges and new research facilities.

Later this year, the ministry of education will launch a blueprint for a 'double excellence initiative' to drive China's universities and academic programmes towards world-class standards, such as by assembling high-quality teams. It is likely that the evaluation for universities will change to reward the achievement of a few top-quality departments rather than many average ones — with similar goals to the UK Research Excellence Framework. Many universities are adjusting their academic structures and realigning leading researchers in anticipation.

### REINFORCE INTEGRITY

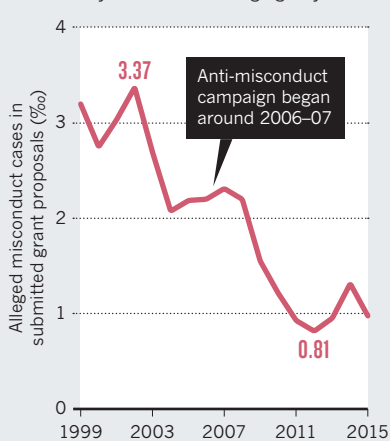
China is enduring a long march to research integrity<sup>2</sup>. The United States tops the league table of retracted papers (see retractionwatch.com), partly owing to its formidable quantity of scientific publications, but retractions from China are growing. The countries have taken slightly different educational approaches to reinforcing integrity<sup>3</sup>. In China, research misconduct tends to be portrayed in black-and-white terms — scientists are either on the moral high ground or cast into the ethical abyss. In the United States, educators analyse grey areas by discussing case studies with early-career researchers in class. Both countries can learn from each other.

Research misconduct in China is driven by several forces<sup>2,3</sup>. These include competition (owing to the rapid expansion of researcher numbers) as well as assessment criteria — the need to publish in international journals encourages the use of language services or ghostwriters, and quantification encourages research outputs to be split up and published separately (known as salami-slicing). Other drivers include strengthening of ethical values such as animal rights, and insufficient provision of ethical codes in areas such as genetics and big data.

For the past decade, the China Association for Science and Technology, the education ministry and the NSFC have run a well-publicized anti-misconduct campaign in the Chinese scientific community. It is having results (see 'Misconduct allegations fall'). Most research institutions now have procedures in place to tackle suspected or confirmed ethical breaches, and a zero-tolerance policy has been enforced in some, such as Zhejiang University in Hangzhou (see

### MISCONDUCT ALLEGATIONS FALL

China is on a long march to research integrity, as shown by data from funding agency NSFC.



SOURCE: NSFC



China's DAMPE satellite launched in December 2015; others are scheduled to follow this year.

XINHUA/REX/SHUTTERSTOCK

*Nature* **481**, 134–136; 2012). The proportion of allegations of misconduct is declining even though more attention is being given to actively detecting cases. Similarities between submitted proposals and published dissertations are also going down. The culture of new researchers is changing from 'why not cheat' to 'it is not worth getting caught'.

To go further requires changes on three fronts: attitude, structure and methodology.

A change in attitude — from covering up misconduct to exposing it — is essential. The NSFC is implementing a similarity check for submitted grant proposals and now publishes an annual press release detailing notorious misconduct cases. It is also investigating cases of ghostwriting and reviewing. Since 2000, we have evolved our policies on information handling, from guarding review confidentiality to transparency in research evaluation. Each panel is required to monitor the healthy conduct of review in its discipline, by voting on the fairness of their fellow panellists' judgements, for example.

Structural changes within institutions are crucial to separate administrative and academic powers and prevent corruption. For instance, the NSFC has exercised various practices that might be applied in other funding agencies. Agency administrators are no longer involved in academic reviews. NSFC staff members are only authorized to access information that is relevant to their duties, and an independent council of senior academics has been set up to counterbalance the administrators. In many institutions, external advisers are now used to avoid conflicts of interest, academic committees are being given more power, and committees have been formed to safeguard research and clinical ethics.

Methodology changes can remove the soil that nourishes research misconduct. A nationwide campaign against overly quantified measures of research is under way. Caps

on human-resource costs are being lifted. A streamlined funding architecture needs to be achieved, which reduces fragmentation in grant sources and mandates that all grant reviews are conducted by professional institutions selected by a joint committee rather than by administrators. We also need to use more external reviews by international peers, and account for indirect research costs.

### PRIORITY TOPICS

Which areas of basic science look most promising to develop in the next five years?

The NSFC's plan for 2016–20 includes a list of areas, breakthroughs and interdisciplinary hotspots in which China could deliver fast. Examples are the 'Langlands programme' for mathematics that links number theory, geometry, analysis and theoretical physics, such as at the Academy of Mathematics and Systems Science, CAS; and the deep underground Earth-physics laboratory near Jinping, Sichuan, that might detect dark matter. There is also the Five-hundred-meter Aperture Spherical Telescope (FAST) in Guizhou, southwest China, which is due to be completed in September; and 24 scientific satellites planned for the next 5 years (4 of which are due to launch this year) that will advance astrophysics, cosmology and Earth sciences.

Other promising areas and institutes include molecular chemistry and quantum catalysis for chemistry, which is a focus of the Dalian Institute of Chemical Physics at CAS; quantum computing for information science at the University of Science and Technology of China in Hefei; and neural circuits and brain science in Shanghai's biomedical sciences and innovation complex. The National Center for Protein Sciences (the PHOENIX Center) in Beijing and Shanghai is focusing on proteomics; and teams are working on gene editing, molecular approaches to cancer, and infectious diseases. A multidisciplinary

effort is needed to stimulate the country's 'green science' — Earth, ocean and environmental science.

### MUCH TO DO

For basic science innovations to benefit the economy, the full chain of development — from the initial research to technology, products and the market — must be nurtured. Not all research will bear fruit beyond the lab; some is curiosity-driven. But where possible, new knowledge should either be turned into technology or translated from one field to others.

The NSFC is feeding the source. The rest of the chain is being encouraged by the Ministry of Science and Technology, through its major National Initiatives for Technology and Engineering (16 of which run to 2020 and 15 that extend to 2030) and National Key Research Projects (36 launched this year). These programmes link researchers, developers and venture capitalists. Examples include addressing air pollution, increasing the use of low-carbon energy in chemical engineering and deep-sea stations for ocean exploration.

Barriers between research and commercial development are being dismantled by new policies. These include the recent revisions of knowledge-transfer laws, which assign the benefits of public-funded projects to the researchers and their institutions (similar to the US Bayh–Dole Act). Researchers thus gain incentives of fame and wealth.

In summary, four issues need attention. First we must incentivize, not discourage, Chinese scientists for making big scientific breakthroughs. These take time and endurance, as the recent detection of gravitational waves illustrates. Areas such as basic physics and astronomy need master plans for long-term development.

Second, we must develop and adopt an assessment strategy using appropriate metrics for evaluating merit.

Third, we must create a healthy, congenial academic ecology. We should let researchers spend time on research, rather than overload them with paperwork or leave them to fend off allegations and slog over grant finances.

Finally, we must devise a business model for China that identifies and cultivates applicable research findings. There are many miles to go before we rest. ■

**Wei Yang** is a professor of engineering mechanics at Zhejiang University, Hangzhou, China. He is president of the National Natural Science Foundation of China in Beijing.  
e-mail: yangwei@nsfc.gov.cn

1. Shi, Y. & Rao Y. *Science* **329**, 1128 (2010).
2. Yang, W. *Science* **342**, 1019 (2013).
3. Frankel, M. S., Leshner, A. I. & Yang, W. in *Handbook of Academic Integrity* (ed. Bretag, T.) 847–866 (Springer, 2016).



John Maynard Keynes (right, with Henry Morgenthau) laid the foundations for the GDP metric.



**The Great Invention: The Story of GDP and the Making (and Unmaking) of the Modern World**  
EHSAN MASOOD  
Pegasus: 2016.

free online services, because the relevant statistics are not collected or do not fit easily into existing categories. There is even a mini-boom in books about economic statistics. A decisive coalition is shaping up in favour of moving away from GDP. The question is what to use instead.

In *The Great Invention*, Ehsan Masood, editor of policy peri-

odicals *Research Europe* and *Research Fortnight*, argues for an improved GDP. Into this single metric for economic-activity indicators — defined as the monetary value of all goods and services produced in a country — he would combine environmental impacts and human well-being. His book traces the history of GDP since its creation, as well as the calls for alternatives, mainly from environmentalists. Masood agrees with the sentiment of suggestions to use ‘dashboards’ that incorporate other economic data and supplementary indicators, but he concludes that GDP matters. As he writes of countries that adopted it: “The act of measuring their economies would ultimately determine how their economies would be managed.” And it matters despite, or because of, its flaws. GDP is too entrenched to be successfully replaced, he finds; instead, it needs radical reform.

#### FORMATIVE FACTORS

GDP began, as Masood notes, as an aggregate measure when the need arose for governments to manage economies during the Depression in the 1930s and the Second World War. Pioneers of the statistics involved, such as US economist and Nobel laureate Simon Kuznets, intended to create a metric to meaningfully capture a society’s economic welfare.

There were other formative factors at work. One was the need to avoid suggesting that the war effort was reducing welfare. Another was the thinking of influential British economist John Maynard Keynes, as set out in his 1936 *The General Theory of Employment, Interest and Money*. Keynes theorized that raising aggregate demand or total spending in the economy through government expenditure can avoid the sort of mass unemployment that was seen in the Depression by stimulating growth and improving stability.

**NATURE.COM**  
For more on science  
in culture see:  
[nature.com/  
booksandarts](http://nature.com/booksandarts)

He and his supporters were determined to make the new metric serve that government role by defining

#### ECONOMICS

# GDP in the dock

**Diane Coyle** savours a history of the long-standing economic measure and possible alternatives.

Since its invention during the Second World War as a thermometer of economic health, gross domestic product (GDP) has become a familiar incantation in claims and counter-claims about the well-being of nations. Some environmentalists and feminists were early critics, but until recent decades, few others questioned it.

Now, campaigners ranging from left-wing Nobel-prizewinning economist Joseph Stiglitz to the free-market *Economist* magazine want to replace GDP with direct measurement of human well-being. The technology industry has joined them, bemoaning the failure of GDP to account properly for digital technologies, including

federal spending as a key component of the equation, along with consumer spending and investment. Thus, GDP was born as a transatlantic effort, led by Keynes's assistants in the UK treasury, Richard Stone and James Meade. By the end of the 1940s, it was standardized through the United Nations, and the same international process is in place today.

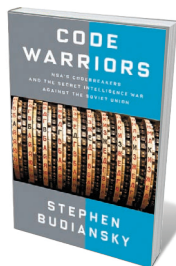
Masood covers decades of challenges to GDP conventions that make for a fascinating institutional and human story. Those seeking an alternative included UN official Maurice Strong, a key figure in the 1972 UN Conference on the Human Environment and the 1992 Earth Summit. Other critics were Italian industrialist Aurelio Peccei and British civil servant Alexander King, who together established think tank the Club

**"The act of measuring economies would ultimately determine how economies would be managed."**

of Rome and published the influential report *The Limits To Growth* (Universe, 1972). Environmentalists, and the officials whom they influenced, were swift to point out that GDP doesn't

take into account how natural assets are depleted to generate current income and consumption. Several proposals for new models have underlined the need to account for the environment; Masood (an erstwhile *Nature* journalist) praises a 1997 paper on these proposals co-authored by economist Robert Costanza (R. Costanza *et al. Nature* **387**, 253–260; 1997). Economists James Tobin and William ▶

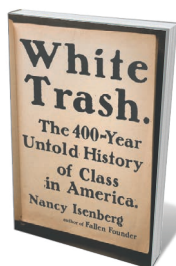
## Books in brief



### Code Warriors

Stephen Budiansky KNOFF (2016)

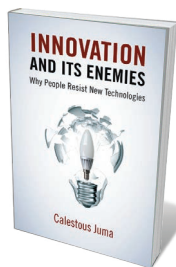
Code fiend and writer Stephen Budiansky's history of the US National Security Agency (NSA) and its intelligence battles with the Soviet Union opens in 2013, as whistle-blower Edward Snowden enacts his long-planned exposé. In a narrative laced with cryptanalysis, Budiansky then tacks back through the NSA's turbulent history, from the "almost insane logical disconnects" of the cold war stand-off to the fall of the Berlin Wall. This is a balanced, authoritative portrait of an institution in which brilliant innovation in mathematics, computing and technology has coexisted with gross invasions of societal privacy.



### White Trash: The 400-Year Untold History of Class in America

Nancy Isenberg VIKING (2016)

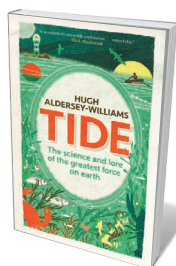
Crackers, clay-eaters and "poor white trash": the white US underclass has endured crass labelling from colonial times. That marginalization begs vast questions about US democracy, argues historian Nancy Isenberg. Her powerful social and cultural history uncovers new facets of known stories, from class conflict in the American Civil War to the sterilization of destitute whites by interwar eugenicists (V. Nourse *Nature* **530**, 418; 2016). At once brutal and enlightening, this is the chronicle of a dispossessed people caught in rural stasis, and the social and political forces that keep them there.



### Innovation and Its Enemies: Why People Resist New Technologies

Calestous Juma OXFORD UNIVERSITY PRESS (2016)

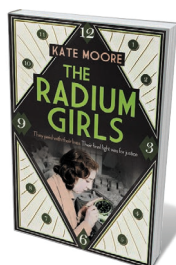
From smart grids to new commodities, innovation disrupts by default — and if it is truly transformative, can trigger controversy and policy headaches. Sustainable-technologies expert Calestous Juma explores those tensions in this original study. He follows coffee from Ethiopia through Europe as it is embraced and denounced, shaping economies, technologies and industries. He looks at the advent of electricity and transgenic crops. For the pace of innovation and institutional change to synchronize, he concludes, both nimble leadership and rigorous, respectful public education must be brought into play.



### Tide: The Science and Lore of the Greatest Force on Earth

Hugh Aldersey-Williams VIKING (2016)

More than 40% of humanity lives within 150 kilometres of a coast, yet a clear understanding of tides — that oceanic phenomenon driven by the gravitational lock of Earth and Moon — is rare. Science writer Hugh Aldersey-Williams's corrective meshes a history of the science (by way of Aristotle, Galileo and Isaac Newton, among others) with tide-related technologies and tidally sculpted events. It's an eloquent ebb and flow, from observations of a 13-hour tidal cycle in a Norfolk salt marsh to passages on the legendary maelstroms of Nova Scotia and California's body-surfing fish, the grunion.



### The Radium Girls

Kate Moore SIMON & SCHUSTER (2016)

In the 1910s, radium was marketed as a cure-all, incorporated into drinking water, cosmetics and even jockstraps. Kate Moore's harrowing chronicle traces how a number of young US women, hired to paint military timepieces with radium-laced paint, paid the price: many succumbed to radiation poisoning and died hideous deaths. Ultimately, the landmark case won by five of them inspired globally important research into radiation and its impacts — including longitudinal studies with survivors. [Barbara Kiser](#)



Amartya Sen created the UN Human Development Index with Mahbub Ul Haq.

THE AGE/GETTY



UN PHOTO/YUTAKA NAGATA

Maurice Strong (front left, at the 1972 UN Conference on the Human Environment) sought an alternative to the GDP.

► Nordhaus also took into account environmental costs — and the value of work in the home — in their 1972 proposal for a metric called the Measure of Economic Welfare.

Another challenge to convention came from Mahbub Ul Haq and Amartya Sen, who in 1990 created the now widely used UN Human Development Index (HDI), which includes factors such as life expectancy and education.

More social scientists are now exploring the definition and use of economic statistics. There is also policy interest alongside the scholarly debate. In 2008, then-French president Nicolas Sarkozy set up a commission led by Sen, Stiglitz and fellow economist Jean-Paul Fitoussi to investigate the measurement of economic well-being. And UK economist Charles Bean's 2016 *Independent Review of UK Economic Statistics* (see [go.nature.com/1tvadaj](http://go.nature.com/1tvadaj)) raises fundamental questions about GDP's viability in a modern economy, for example concerning its mismeasurement of digital activity.

#### ONE OR MANY

The balance of opinion in economics currently favours supplementing GDP with a dashboard that incorporates measures of environmental impacts, health and social indicators, as Costanza neatly summarized in his 1997 article (see also R. Costanza *et al.*

*Nature* 505, 283–285; 2014). Economists are taking considerable interest in the measurement debate, although oddly, Masood claims that the profession is ignoring the issue. His own call for a nuanced metric that factors in natural capital and human well-being sticks to one indicator. He thinks that GDP is so tightly woven into the economic fabric that anything more complicated than a single number will put politicians and the media off. The solution, as he sees it, is “to value the things that matter and then incorporate this value into the GDP accounts”.

But single-indicator alternatives to GDP have a serious drawback. They hide relative valuations of their components, whereas GDP makes these explicit because it uses market prices. For instance, Martin Ravallion, former director of research at the World Bank, notes that the HDI implicitly values poor lives much less than rich ones. Because income and human life expectancy are combined into one index, there is an implied value of just US\$0.51 for an extra year of life in Zimbabwe, compared

*“GDP is so tightly woven into the economic fabric that anything more complicated than a single number will put politicians and the media off.”*

to several thousand dollars in rich countries (see M. Ravallion *Troubling Tradeoffs in the Human Development Index* <http://doi.org/d8d2cr>; World Bank, 2010). This flaw could be corrected, but the point is that any single index internalizes such trade-offs.

The debate over whether to use a dashboard or a single indicator is unresolved. Interest among economists, other social scientists and environmentalists has climbed in recent years, but there is much to research and discuss on how best to measure economic welfare, taking into account sustainability and the quality of life, before a new international standard is defined and adopted.

In hindsight, the original debate about GDP looks more compressed than it really was — some economists were still disputing it into the 1950s. A new shift will take just as long, but it is definitely under way. And about time too, for the reasons that *The Great Invention* explains so clearly. ■

**Diane Coyle** is a professor of economics at the University of Manchester, UK, and author of *GDP: A Brief but Affectionate History*. She is also a member of the UK government's Natural Capital Committee and an Office for National Statistics Fellow.  
e-mail: [diane.coyle@manchester.ac.uk](mailto:diane.coyle@manchester.ac.uk)

federal spending as a key component of the equation, along with consumer spending and investment. Thus, GDP was born as a transatlantic effort, led by Keynes's assistants in the UK treasury, Richard Stone and James Meade. By the end of the 1940s, it was standardized through the United Nations, and the same international process is in place today.

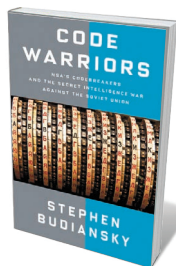
Masood covers decades of challenges to GDP conventions that make for a fascinating institutional and human story. Those seeking an alternative included UN official Maurice Strong, a key figure in the 1972 UN Conference on the Human Environment and the 1992 Earth Summit. Other critics were Italian industrialist Aurelio Peccei and British civil servant Alexander King, who together established think tank the Club

**"The act of measuring economies would ultimately determine how economies would be managed."**

of Rome and published the influential report *The Limits To Growth* (Universe, 1972). Environmentalists, and the officials whom they influenced, were swift to point out that GDP doesn't

take into account how natural assets are depleted to generate current income and consumption. Several proposals for new models have underlined the need to account for the environment; Masood (an erstwhile *Nature* journalist) praises a 1997 paper on these proposals co-authored by economist Robert Costanza (R. Costanza *et al. Nature* **387**, 253–260; 1997). Economists James Tobin and William ▶

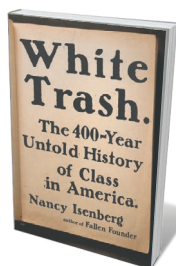
## Books in brief



### Code Warriors

Stephen Budiansky KNOFF (2016)

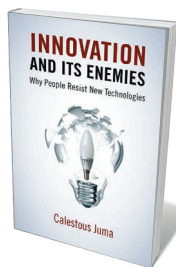
Code fiend and writer Stephen Budiansky's history of the US National Security Agency (NSA) and its intelligence battles with the Soviet Union opens in 2013, as whistle-blower Edward Snowden enacts his long-planned exposé. In a narrative laced with cryptanalysis, Budiansky then tacks back through the NSA's turbulent history, from the "almost insane logical disconnects" of the cold war stand-off to the fall of the Berlin Wall. This is a balanced, authoritative portrait of an institution in which brilliant innovation in mathematics, computing and technology has coexisted with gross invasions of societal privacy.



### White Trash: The 400-Year Untold History of Class in America

Nancy Isenberg VIKING (2016)

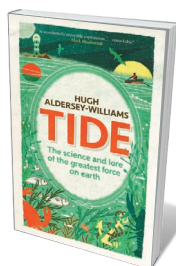
Crackers, clay-eaters and "poor white trash": the white US underclass has endured crass labelling from colonial times. That marginalization begs vast questions about US democracy, argues historian Nancy Isenberg. Her powerful social and cultural history uncovers new facets of known stories, from class conflict in the American Civil War to the sterilization of destitute whites by interwar eugenicists (V. Nourse *Nature* **530**, 418; 2016). At once brutal and enlightening, this is the chronicle of a dispossessed people caught in rural stasis, and the social and political forces that keep them there.



### Innovation and Its Enemies: Why People Resist New Technologies

Calestous Juma OXFORD UNIVERSITY PRESS (2016)

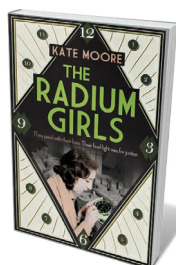
From smart grids to new commodities, innovation disrupts by default — and if it is truly transformative, can trigger controversy and policy headaches. Sustainable-technologies expert Calestous Juma explores those tensions in this original study. He follows coffee from Ethiopia through Europe as it is embraced and denounced, shaping economies, technologies and industries. He looks at the advent of electricity and transgenic crops. For the pace of innovation and institutional change to synchronize, he concludes, both nimble leadership and rigorous, respectful public education must be brought into play.



### Tide: The Science and Lore of the Greatest Force on Earth

Hugh Aldersey-Williams VIKING (2016)

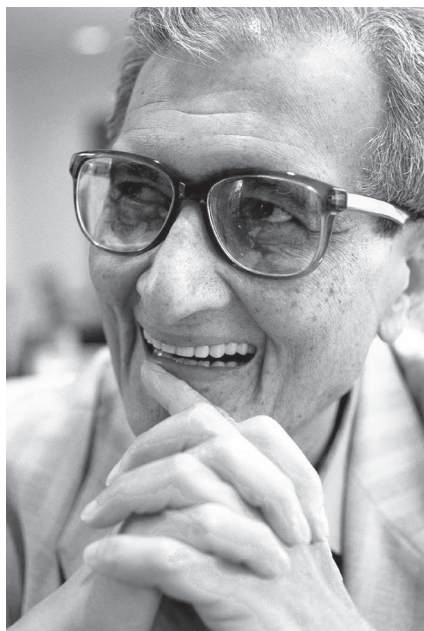
More than 40% of humanity lives within 150 kilometres of a coast, yet a clear understanding of tides — that oceanic phenomenon driven by the gravitational lock of Earth and Moon — is rare. Science writer Hugh Aldersey-Williams's corrective meshes a history of the science (by way of Aristotle, Galileo and Isaac Newton, among others) with tide-related technologies and tidally sculpted events. It's an eloquent ebb and flow, from observations of a 13-hour tidal cycle in a Norfolk salt marsh to passages on the legendary maelstroms of Nova Scotia and California's body-surfing fish, the grunion.



### The Radium Girls

Kate Moore SIMON & SCHUSTER (2016)

In the 1910s, radium was marketed as a cure-all, incorporated into drinking water, cosmetics and even jockstraps. Kate Moore's harrowing chronicle traces how a number of young US women, hired to paint military timepieces with radium-laced paint, paid the price: many succumbed to radiation poisoning and died hideous deaths. Ultimately, the landmark case won by five of them inspired globally important research into radiation and its impacts — including longitudinal studies with survivors. [Barbara Kiser](#)



Amartya Sen created the UN Human Development Index with Mahbub Ul Haq.

THE AGE/GETTY

# Correspondence

## Commit to equity for women researchers

Heads of research agencies from nearly 50 countries — large and small, with developed and emerging economies — adopted a Statement of Principles and Actions Promoting the Equality and Status of Women in Research at the Global Research Council's fifth annual meeting last month in New Delhi (see [go.nature.com/1yqtyg](http://go.nature.com/1yqtyg)).

According to a report commissioned by the Science and Engineering Research Board of India and Research Councils UK, which hosted the meeting, women make up only 11% of full science and engineering professors in the European Union, less than 25% of academics in Asia and less than 5% of researchers in some Middle Eastern countries (see [go.nature.com/1uywmg](http://go.nature.com/1uywmg)). The report echoes statistics from the US National Science Foundation, of which I am director (see [go.nature.com/1rpvmrk](http://go.nature.com/1rpvmrk) for the Science and Engineering Indicators).

At the meeting, we gained greater awareness of long-standing historical obstacles to women's participation in certain fields, and of the importance of including gender considerations in research design and outcome analysis. Each of us came away with a firmer idea of the opportunities to lead within our jurisdictions, and in a wider policy context.

The national research heads agreed to “expect and encourage improved equality and diversity policies and practices” within their respective research provinces, and recommended a list of actions. These included diversity training, recognizing unconscious bias, implementing family-friendly policies and creating pathways for women to rise to leadership positions. We agreed to collect follow-up data and make them available for comparative analysis.

Only by supporting the best talent — wherever it hails from

— can we truly encourage and support research with the greatest academic, economic and societal impacts. Ensuring global equity for women in research requires that we each make a personal commitment to action.

**France A. Córdova** *National Science Foundation, USA.*  
[acollins@nsf.gov](mailto:acollins@nsf.gov)

## Don't bank African rhinos in Australia

The Australian Rhino Project (see [go.nature.com/28c8s29](http://go.nature.com/28c8s29)) aims to move 80 rhinoceroses from South Africa to Australia by 2019 as conservation ‘insurance’ against the poaching epidemic — at a cost of about US\$3.5 million. The first 6 will go this year. In our view, this project is diverting funds and public interest away from the actions necessary to conserve the animals in Africa.

The scheme is supported by the South African and Australian governments, academic institutions in Australia, and corporations and conservation-management organizations. Its cost equates to more than the anti-poaching budget of South African National Parks for 2015. We suggest that this money would be better spent on local, on-the-ground action in South Africa or on education programmes in Asia to reduce demand for rhino horn.

Africa's rhinos are not even the highest priority in pachyderm conservation, particularly because only white rhinos from private collections are to be moved. The global estimated populations of white and black rhinos are 20,170 and 4,880, respectively — still further from extinction than Indian (2,575), Sumatran (275) and Javan (60) rhinos.

We feel that the project has echoes of colonial times, when African resources were exploited. Taking biodiversity assets such as rhinos for ‘safe keeping’ in the West seems to us as patronizing and disempowering as the theft of cultural artefacts.

**Matt W. Hayward\*** *Bangor University, UK.*

[m.hayward@bangor.ac.uk](mailto:m.hayward@bangor.ac.uk)

*\*On behalf of 4 correspondents (see [go.nature.com/1w32n9q](http://go.nature.com/1w32n9q) for full list).*

## Freelance scientists need EU for support

As ‘freelance’ scientists, we undertake research jointly with academic institutions and provide Earth-science modelling services for clients — an alternative career path that European Union funding enables us to pursue. If the United Kingdom chooses to leave the EU after this week's referendum, small private research organizations and independent researchers could be doomed.

Independent researchers cannot apply for funding from UK research councils. Private research organizations need demonstrable in-house research capacity and a minimum of ten researchers. These eligibility criteria are at odds with those of the UK arts councils and the European Commission, which consider proposals from anyone with a track record in their discipline.

With 88% of UK postdocs never securing a tenured position (*The Scientific Century: Securing our Future Prosperity*; Royal Society, 2010), these requirements need to be relaxed (see also *Nature* **520**, 144–147; 2015). Entrepreneurial young scientists could then continue their research without the backing of a university.

For the United Kingdom to maintain its competitive edge, funding bodies need to recognize that the research landscape is changing. In this era of digital connectivity, scientists can still be embedded in the research community while working outside traditional research organizations.

**Cécile B. Ménard, Melody Sandells** *CORES Science and Engineering, Burnopfield, Newcastle upon Tyne, UK.*  
[cecile.menard@coresscience.co.uk](mailto:cecile.menard@coresscience.co.uk)

## Carry on celebrating Mendel's legacy

I disagree with Gregory Radick's strategy for teaching modern genetics (*Nature* **533**, 293; 2016). In my view, we should not discard the legacies of Gregor Mendel, William Bateson, Walter Sutton, Thomas Hunt Morgan and their ilk, whose beautiful science continues to provide the best explanations for inheritance.

I teach basic genetics to veterinary students, who learn the laws of inheritance without any historical context, and to biology students, who learn the scientific method and how it influenced the development of genetic concepts. The biologists revisit hypotheses proposed to account for the same observations — such as Bateson's and W.F.R. Weldon's contrasting views of inheritance. They come to understand that Mendel's hypothesis of hereditary units (‘alleles’) explains the data better. They learn that theories and hypotheses are not immutable, that science is incomplete, and that every discovery stimulates new questions.

With the Boveri–Sutton chromosome theory, it became clear that Mendelian inheritance is indeed the core of genetics. It underpins association-mapping studies, population genetics and clinical genetics. Such new information continues to corroborate Mendel's hypothesis of inheritance. There is no need to remove Mendel from his honorary position in the genetics curriculum to spark creative science.

**Tatiana T. Torres** *Institute of Biosciences, University of São Paulo, Brazil.*  
[ttorres@ib.usp.br](mailto:ttorres@ib.usp.br)

### CONTRIBUTIONS

Correspondence may be sent to [correspondence@nature.com](mailto:correspondence@nature.com) after consulting [go.nature.com/cmchno](http://go.nature.com/cmchno).

## ASTROPHYSICS

## Recipe for a black-hole merger

The detection of a gravitational wave was a historic event that heralded a new phase of astronomy. A numerical model of the Universe now allows researchers to tell the story of the black-hole system that caused the wave. [SEE LETTER P.512](#)

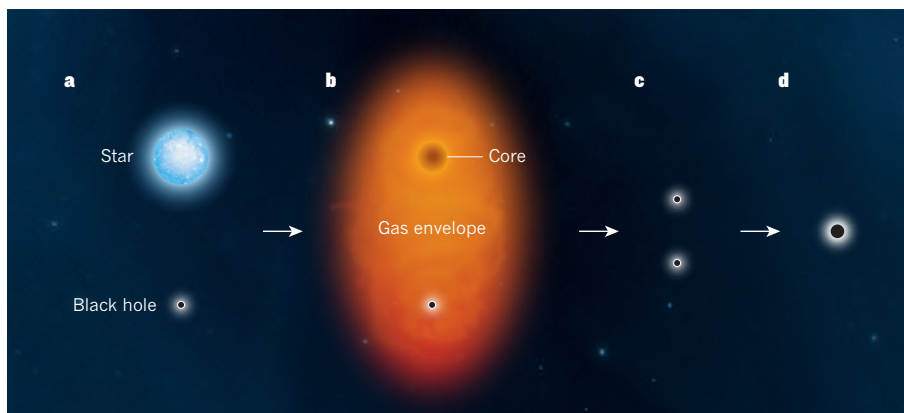
J. J. ELDRIDGE

The first gravitational-wave source was detected on 14 September 2015. What surprised some was that the signal came from the merger of two black holes, each about 30 times the mass of the Sun. Now, on page 512, Belczynski *et al.*<sup>1</sup> not only show that such a system can arise naturally from our understanding of how the stars in binary systems interact, but also unlock the history of the black holes from their birth as two massive stars.

Other groups<sup>2–4</sup> have sought to characterize the source of the gravitational wave (now known as GW 150914), but what makes Belczynski and colleagues' work stand out is that they have created a numerical model of the Universe that allows every phase of the evolution of binary stars to be followed, from the birth of the Universe to the present. This enabled them to search through the list of observable black-hole binaries to find those that match the parameters of the gravitational-wave source. They then tracked back each candidate source's evolution to estimate the relative probability that the source could have caused the event, and thus identify which was the most likely.

The authors conclude that the black holes probably started off as two stars that had masses 40 to 100 times that of the Sun and were born about 2 billion years after the Big Bang. These stars turned into black holes after a further 5 million years, and merged 10.3 billion years after that, emitting the gravitational-wave signal that was detected 1.2 billion years later (Fig. 1). Other scenarios are possible, but less likely.

The black holes were monsters, and the results show that their progenitor stars would have been some of the brightest and most massive in the Universe. If the proposed age of the stars' formation is correct, then they might have contributed to the reionization of the Universe — one of the key events in the Universe's evolution. It is also likely that the stars were relatively pure in composition: they consisted mostly of hydrogen and helium, and contained less than 10% of the heavy elements (such as carbon, oxygen and iron) that pollute our Sun. This indicates that the stars



**Figure 1 | Key interactions that led to a black-hole merger.** Belczynski *et al.*<sup>1</sup> used numerical models of the Universe to unlock the history of the black-hole binary system that caused the gravitational wave reported in 2015. **a**, They propose that one of the two stars in the progenitor binary system exploded as a supernova (not shown), forming a black hole. **b**, That was engulfed by the second star as it evolved and expanded, generating a system in which the two objects shared the same gas envelope. **c**, Interaction between the two objects gradually decreased the distance between them, and the second star formed a black hole. **d**, The two black holes continued to get closer by radiating gravitational waves, eventually merging and generating gravitational waves strong enough to be detected.

would have been in a small dwarf galaxy, rather than a large spiral galaxy, such as our own Milky Way.

This study is important for two reasons. First, GW 150914 provides an exciting test for stellar evolution theory. Previously, core-collapse supernovae represented the latest stage of a star's life that could be used to constrain the nature of the progenitor stars<sup>5</sup>. Belczynski *et al.* have gone beyond that to the final event that occurs within a stellar binary that has already survived two supernovae. Their work, therefore, places firm constraints on stellar evolution and on how stars die in supernovae. Second, it provides a new way to measure the accuracy of models of star formation and cosmic evolution throughout the history of the Universe.

There are, of course, caveats and assumptions that add uncertainty to Belczynski and colleagues' model. One uncertainty is how massive the black hole formed by a star can be; this is determined by how explosive the black-hole-forming supernovae are. The explosive nature of massive stars is a hot topic of research, with some evidence<sup>6,7</sup> suggesting that black holes can form directly from stars without a supernova, which is what Belczynski and

colleagues assume. But stars might also form black holes and explode. In binary systems, this would affect the nature of the final black-hole system, and the time taken for the black holes to merge.

Another uncertainty involves an intermediate phase of binary-star evolution. As stars in binaries evolve, their radii increase, sometimes growing to the size of their orbit so that they get in each other's way — this is called the common-envelope phase. Typically, the star that grows first loses its outer envelope of gas, leaving a small, hot core that eventually forms a black hole. The binary's orbit decreases in size during this process. During the merger of two black holes, the closer the two objects were when they formed, the sooner they will merge. But researchers do not know how much the orbit can shrink in the common-envelope phase despite decades of work dedicated to finding an answer<sup>8</sup>.

Future gravitational-wave signals may help astrophysicists to constrain both uncertainties, but for now, Belczynski *et al.* generate an 'optimistic' and a 'pessimistic' model Universe, to assess the highest and lowest possible rates of black-hole mergers. They demonstrate that systems that would form black-hole binaries

of the sort that generated GW 150914 would form in both models, and that the rate of black-hole mergers in the Universe matches that inferred from the gravitational-wave detection. The authors also suggest that rotation of stars about their own axes is not required to explain most gravitational-wave sources, but it has been suggested that such rotation could increase the number of black-hole mergers<sup>9</sup>. Nevertheless, there is still more work to be undertaken, and more physics to include in the models.

Belczynski and colleagues' study is

tremendously exciting because it examines the effects of a new constraint on how stars and the Universe evolve, identified by GW 150914. With each gravitational-wave signal detected we'll learn something new. And with rumours that more events will be announced soon, we may not have to wait too long for the next lesson. ■

**J. J. Eldridge** is in the Department of Physics, University of Auckland, Private Bag 92019, Auckland, New Zealand.  
e-mail: j.eldridge@auckland.ac.nz

1. Belczynski, K., Holz, D. E., Bulik, T. & O'Shaughnessy, R. *Nature* **534**, 512–515 (2016).
2. Eldridge, J. J. & Stanway, E. R. <http://arxiv.org/abs/1602.03790> (2016).
3. Hartwig, T. et al. *Mon. Not. R. Astron. Soc.* **460**, L74–L78 (2016).
4. Lipunov, V. M. et al. <http://arxiv.org/abs/1605.01604> (2016).
5. Smartt, S. J. *Publ. Astron. Soc. Aust.* **32**, e016 (2015).
6. Sukhbold, T., Ertl, T., Woosley, S. E., Brown, J. M. & Janka, H.-T. *Astrophys. J.* **821**, 38 (2016).
7. Gerke, J. R., Kochanek, C. S. & Stanek, K. Z. *Mon. Not. R. Astron. Soc.* **450**, 3289–3305 (2015).
8. Ivanova, N. et al. *Astron. Astrophys. Rev.* **21**, 59 (2013).
9. de Mink, S. E. & Mandel, I. *Mon. Not. R. Astron. Soc.* <http://dx.doi.org/10.1093/mnras/stw1219> (2016).

## CELL BIOLOGY

# Membrane kiss mediates hormone secretion

Communication between cells relies on hormone release from secretory granules, but how these vesicles fuse with cell membranes is unclear. An imaging study provides *in vivo* evidence for a stable intermediate fusion step. **SEE LETTER P.548**

TOLGA SOYKAN & VOLKER HAUCKE

Cells communicate with each other by secreting small messenger molecules such as hormones or growth factors, many of which are stored in vesicles called secretory granules. To release these messengers to the cell exterior, secretory granules fuse with the cell membrane through a process called exocytosis. On page 548, Zhao *et al.*<sup>1</sup> show that exocytosis occurs through the reversible formation of a hemi-fused intermediate, in which only one of the two leaflets

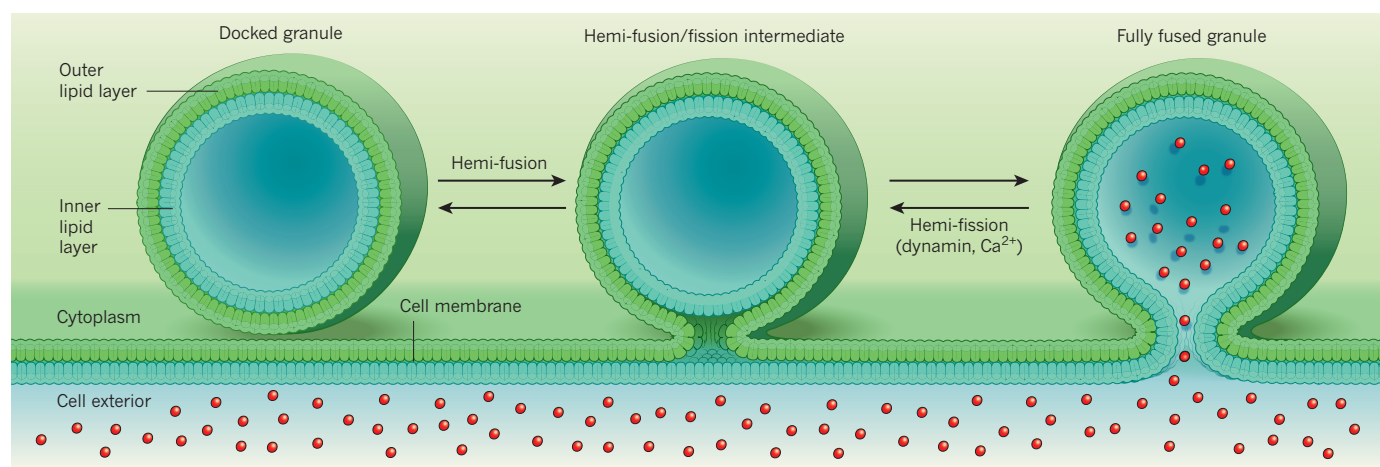
(lipid layers) of the cell membrane has merged with the secretory granule's membrane. These results answer the long-standing question of whether membrane fusion involves a hemi-fused intermediate and also provide *in vivo* evidence for the 'kiss-and-run' model of secretory-granule exocytosis.

Examples of hormones that are released by secretory granules include insulin in pancreatic  $\beta$ -cells and adrenaline, found mainly in chromaffin cells of the adrenal glands. Hormone secretion requires the granules to dock with the cell membrane and then partially or completely

merge with it<sup>2,3</sup>, allowing hormone release into the extracellular space. However, the regulatory mechanism that underlies this type of membrane fusion<sup>3</sup>, and the nature of the intermediates involved, have long been debated<sup>4,5</sup>.

Zhao *et al.* used a high-resolution optical-microscopy approach<sup>6</sup> to observe membrane fusion directly in living chromaffin cells in real time and in three dimensions. To trigger secretory-granule release, the chromaffin cells were electrically stimulated while being bathed in a cell-impermeable fluorescent marker dye. The granule's lipid layer maintains a tight seal, so that the dye can gain access to the interior of a docked secretory granule only when full fusion of the granule and cell membrane occurs. Use of this dye, along with simultaneous monitoring of the cytoplasmic leaflet of the cell membrane using a lipid-bound fluorescent protein, enabled the authors to analyse the membrane changes that occur during fusion.

As the granules begin to fuse, diffusion of the fluorescent marker protein from the inner leaflet of the cell membrane to the outer leaflet of the granule membrane results in an increase in fluorescence in the granule that



**Figure 1 | A hemi-fused intermediate in membrane fusion.** Zhao *et al.*<sup>1</sup> observed in chromaffin cells that secretory granules that have docked with the cell membrane can undergo reversible hemi-fusion, in which the leaflets of the outer lipid layer of the secretory granule and the cytoplasmic-facing lipid layer of the cell membrane merge. In the authors' experiments, the inner lipid layer of the hemi-fused secretory granule maintains a tight seal that

prevents access of external dye marker molecules (red). When the secretory granules undergo full fusion with the cell membrane, the external dye can enter the secretory granules. Fully fused secretory granules can revert to the hemi-fusion/fission intermediate in a process that depends on binding of the protein dynamin and is regulated by cytoplasmic calcium ( $\text{Ca}^{2+}$ ) levels.

serves as a reporter for changes in the fusing membranes. In more than half of all fusion events, the authors observed a simultaneous influx of fluorescent dye into the secretory-granule lumen and rise in membrane fluorescence on the docked secretory granule. This type of fluorescence change would be expected if secretory granules fuse completely with the cell membrane.

The authors also frequently observed events in which a rise in fluorescence of the cell-membrane marker on the secretory granules preceded dye influx by several seconds, or in which no dye influx into the secretory granule was detected during 40 seconds of observation. It was surprising that so many such events were detectable and that they are stable over many seconds. From these and further control experiments, the authors concluded that these events must correspond to hemi-fusion, thought to be a metastable state in which the outer leaflet of the secretory granule and the cytoplasm-facing leaflet of the cell membrane have merged, while the inner leaflet of the secretory granule and the extracellular-facing leaflet of the cell membrane remain separate<sup>4</sup>.

What might be the advantages of this type of fusion mechanism? Unlike neuronal synaptic vesicles, which release neurotransmitter molecules in an all-or-nothing, 'quantal' fashion, secretory granules can partially secrete their hormone content. Such partial release has been postulated to be mediated by kiss-and-run exocytosis, which is a model for how secretory granules open and close a fusion pore through which molecules pass between the granule and the cell membrane<sup>2</sup>. Whether the fusion pore is made of lipid or protein, or both, is not known. A stable hemi-fused intermediate might indicate the existence of a reversible fusion process that would enable partial release during secretory-granule exocytosis, and might underlie fusion-pore opening and closing<sup>2</sup>.

An identical structure to a hemi-fused intermediate could also arise if a fully fused granule underwent fusion-pore closure through fission (the splitting of a membrane into two separate entities). In this context, it would be called a hemi-fission intermediate. To probe whether fusion-pore opening and closing are reversible processes that proceed through a common intermediate, Zhao *et al.* tracked the movement of the fluorescent dye and membrane markers over time. They frequently observed fluorescence dynamics consistent with closure of the fusion pore through a hemi-fission state.

The authors then investigated the role of dynamin, a protein involved in endocytosis — a cell-membrane-dependent process in which materials are transported into cells. Dynamin can bind<sup>7</sup> to narrow lipid 'necks' at places where membrane pinching occurs, and Zhao *et al.* found that its depletion or inhibition tipped the balance towards full fusion of both layers of

the granule and cell membranes at the expense of hemi-fusion events. Conversely, the hemi-fused state seemed to be stabilized by a high influx of calcium into the cytoplasm. Overall, the authors' data indicate that secretory-granule fusion and fission are reversible processes, at least in chromaffin cells, with the transition from hemi-fusion to full fusion being counteracted by dynamin and regulated by cytoplasmic calcium (Fig. 1). Consistent with this model, Zhao *et al.* occasionally observed reversible opening, closing and reopening of fusion pores in the same docked secretory granules.

Although hemi-fusion has previously been observed and characterized in reconstituted systems *in vitro*<sup>8–10</sup>, Zhao and colleagues' work is the first demonstration of this process in living cells. The new results indicate that this intermediate fusion state is a physiologically relevant and surprisingly stable intermediate en route to the exocytic release of hormones and related molecules. It has been suggested that hemi-fusion underlies the fusion of yeast membrane-bound structures called vacuoles<sup>11</sup>, and that it is also probably responsible for the delayed fusion pathway in reconstituted vesicles *in vitro*<sup>8</sup>. However, the authors' model of exocytic membrane fusion is difficult to reconcile with the idea that the fusion pore is lined with transmembrane proteins — as has been postulated from mutational analysis of the transmembrane segments of key exocytic proteins<sup>5,12</sup> — because transmembrane proteins span both layers of the membrane and therefore would be excluded from the centre of the hemi-fused intermediate.

Whether a mechanism that involves hemi-fused intermediates operates in neurons to

release neurotransmitters also remains an open question. However, the direct visualization of membrane fusion during neurotransmission presents special challenges, such as the speed of exocytosis, the small size of synaptic vesicles and the complex architecture of neurons in the brain. Finally, the observation that dynamin regulates the partitioning between hemi- and full fusion or fission events lends further support to the idea that membrane fission during endocytosis and other vesicle-budding events proceeds through hemi-fission intermediates<sup>13</sup>. ■

**Tolga Soykan and Volker Haucke** are in the Department of Molecular Pharmacology and Cell Biology, Leibniz Institut für Molekulare Pharmakologie, 13125 Berlin, Germany. e-mails: haucke@fmp-berlin.de; soykan@fmp-berlin.de

1. Zhao, W.-D. *et al.* *Nature* **534**, 548–552 (2016).
2. Albillos, A. *et al.* *Nature* **389**, 509–512 (1997).
3. Südhof, T. C. & Rothman, J. E. *Science* **323**, 474–477 (2009).
4. Chernomordik, L. V. & Kozlov, M. M. *Nature Struct. Mol. Biol.* **15**, 675–683 (2008).
5. Jackson, M. B. & Chapman, E. R. *Nature Struct. Mol. Biol.* **15**, 684–689 (2008).
6. Chiang, H.-C. *et al.* *Nature Commun.* **5**, 3356 (2014).
7. Ferguson, S. M. & De Camilli, P. *Nature Rev. Mol. Cell Biol.* **13**, 75–88 (2012).
8. Diao, J. *et al.* *eLife* **1**, e00109 (2012).
9. Hernandez, J. M. *et al.* *Science* **336**, 1581–1584 (2012).
10. Xu, Y., Zhang, F., Su, Z., McNew, J. A. & Shin, Y.-K. *Nature Struct. Mol. Biol.* **12**, 417–422 (2005).
11. Reese, C., Heise, F. & Mayer, A. *Nature* **436**, 410–414 (2005).
12. Han, X., Wang, C.-T., Bai, J., Chapman, E. R. & Jackson, M. B. *Science* **304**, 289–292 (2004).
13. Mattila, J.-P. *et al.* *Nature* **524**, 109–113 (2015).

This article was published online on 15 June 2016.

## PARTICLE PHYSICS

# Quantum simulation of fundamental physics

**Gauge theories underpin the standard model of particle physics, but are difficult to study using conventional computational methods. An experimental quantum system opens up fresh avenues of investigation. SEE LETTER P.516**

EREZ ZOHAR

There are many questions still to be answered about the standard model of particle physics, which describes the fundamental forces and interactions of nature. On page 516, Martinez *et al.*<sup>1</sup> report a pioneering experiment in which calcium ions that are trapped and controlled by electromagnetic fields form a quantum simulator of elementary particle physics. This is a first experimental step towards the use of quantum simulators to

answer some of those outstanding questions.

Theoretical physics often involves problems that do not have a simple mathematical solution. This quandary is usually overcome using numerical calculations performed by conventional (classical) computers. Some problems, however, cannot be solved by these techniques, and require other methods, especially when direct experimental study is also impossible or difficult.

Physicist Richard Feynman suggested that, to simulate the quantum behaviour of physical

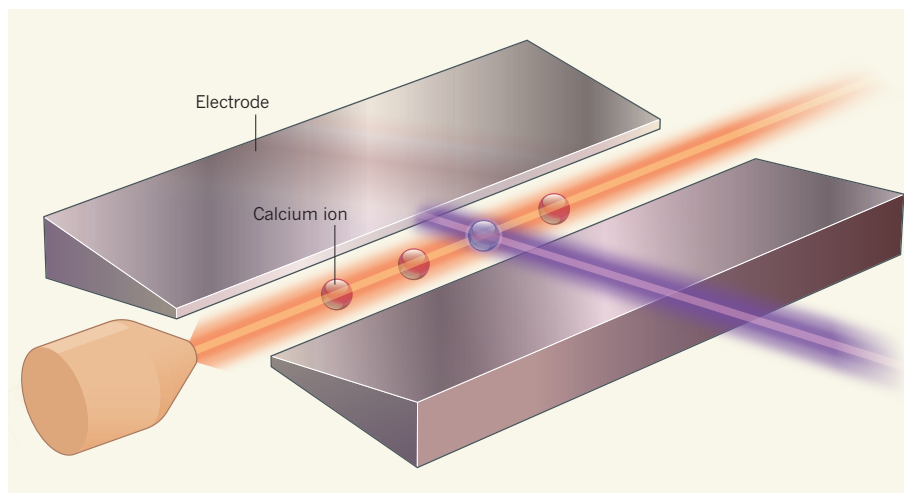
systems, other quantum systems must be used — quantum computers<sup>2</sup>. This concept, called quantum simulation<sup>3</sup>, is beautifully simple. Consider a quantum system, A, that cannot be studied by conventional theoretical and experimental methods, and another quantum system, B, that can be built and controlled with high precision in the laboratory. If the physical components of B, and the interactions between them, mimic and behave like those of A, then B is a quantum simulator of A. Once B is built, tuned and operated, experimental study of it effectively serves as a study of A.

Quantum simulators can be either analog, in which system B simply ‘behaves’ like system A because its dynamics and interactions exactly or approximately map those of A, or digital, in which a sequence of operations acts on the components of B and possibly on some auxiliary elements, generating dynamics that are equivalent to those of A with controlled precision. The simulating systems are often atomic or optical, and have included systems of cold atoms or ions trapped by electromagnetic fields. These have been designed (and some have been built) to simulate many areas of physics, ranging from condensed-matter physics to gravitational effects<sup>3</sup>.

The interactions between elementary particles are a great candidate for quantum simulation. In the standard model of particle physics, such interactions are mediated by vector fields known as gauge fields, thanks to a special type of symmetry called local gauge invariance. Electrons, for example, interact according to the quantum theory of electrodynamics (QED, the simplest gauge theory) through the electromagnetic gauge field. Other fundamental constituents of matter are quarks, whose interactions through the strong force are described by another gauge theory, quantum chromodynamics (QCD).

QCD has several open questions. One is the phenomenon of confinement, in which quarks are bound together by the strong force to form composite particles called hadrons (which include protons and neutrons). The strong force prevents quarks from being isolated experimentally. The theoretical study of confinement is also difficult, and has been a subject of research for decades. A highly successful avenue for studying gauge theories is called lattice gauge theory<sup>4</sup>, but using it for conventional computer simulations is still problematic for the study of several questions.

The quantum simulation of lattice gauge theories has been a rapidly growing area of study over the past few years, and several proposals have been made for how such simulators could be realized<sup>5,6</sup>. These simulators quantitatively map the simulated system — which is typically highly energetic — onto low-energy atomic and optical experimental systems. Martinez *et al.* report the first experimental realization of just such a quantum simulator.



**Figure 1 | Quantum simulation of a gauge theory.** The quantum theory of electrodynamics (QED) is the simplest gauge theory (a type of field theory), and describes how particles of matter such as electrons interact through an electromagnetic field. Martinez *et al.*<sup>1</sup> have used a system of four calcium ions confined by electromagnetic fields (not shown), generated by electrodes, to simulate a one-dimensional model of a variant of this theory known as lattice QED. Each ion serves as a quantum bit that can adopt one of two states, representing the presence or absence of particles, and the interactions between the ions are tailored to represent the gauge-field dynamics. The state of each ion, and the interactions between the ions, can be manipulated using laser beams (orange and purple).

The authors simulated lattice QED in a one-dimensional space (a lattice Schwinger model) using a digital quantum simulator — a tailored quantum computer that consists of four trapped calcium ions controlled and manipulated by lasers (Fig. 1). Two energy levels of each ion form a quantum bit (a qubit), which represents the presence or absence of a particle of matter in the corresponding simulated theory. The gauge field is represented as interactions between the ions that are direct and exotic, yet experimentally implementable. This is achieved using a theoretical transformation available in one dimension that eliminates direct manifestations of the field in the simulated model and allows it to be expressed in terms of matter.

The quantum simulation of complicated gauge theories requires a non-trivial combination of advanced technologies in atomic and optical physics. Martinez and colleagues therefore investigated a small version of a 1D lattice QED model, a relatively simple system that enabled their results to be compared with predictions, but that still demonstrates important features of more-complicated models. The authors’ quantum simulator did indeed reproduce the expected physical behaviour of the simulated model with great accuracy.

In future work, larger systems should be simulated that have a greater number of dimensions (to reveal further non-trivial types of interaction) and involve more-complicated simulated models such as QCD. Quantum simulators for many of these models have already been proposed — both analog<sup>5,6</sup> and digital<sup>7</sup> — for various gauge theories in different dimensions, mostly using cold atoms, but also trapped ions and superconducting qubits.

The experimental requirements and feasibility of these proposals vary, because the simulators use different approaches and involve various simulated models, but they mostly require combinations of existing experimental techniques. Technological developments will help to make such experiments more achievable, even for the simulation of complicated models. As the first quantum simulator of a lattice gauge theory to be built, Martinez and co-workers’ system serves as a beacon that will lead gauge-theory physicists to the promised land of experimental realization.

The authors’ work proves that it is indeed realistic to use quantum-optics techniques to study particle physics and fundamental forces. Further theoretical and experimental advances might enable quantum simulators to solve challenges such as study of the exotic phases of QCD, and to observe new phenomena. More generally, this realization of the great power of quantum simulation reminds us how wonderfully multidisciplinary physics is. ■

**Erez Zohar** is in the Theory Division, Max Planck Institute of Quantum Optics, 85748 Garching, Germany.  
e-mail: [erez.zohar@mpq.mpg.de](mailto:erez.zohar@mpq.mpg.de)

1. Martinez, E. A. *et al.* *Nature* **534**, 516–519 (2016).
2. Feynman, R. P. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
3. Nature Physics Insight: Quantum simulation *Nature Phys.* [www.nature.com/nphys/insight/quantum-simulation](http://www.nature.com/nphys/insight/quantum-simulation) (2012).
4. Wilson, K. G. *Phys. Rev. D* **10**, 2445–2459 (1974).
5. Wiese, U.-J. *Ann. Phys. (Leipz.)* **525**, 777–796 (2013).
6. Zohar, E., Cirac, J. I. & Reznik, B. *Rep. Prog. Phys.* **79**, 014401 (2016).
7. Tagliacozzo, L., Celi, A., Zamora, A. & Lewenstein, M. *Ann. Phys.* **330**, 160–191 (2013).

## EVOLUTION

# Gene regulation in transition

An in-depth analysis of a close relative of animals, *Capsaspora owczarzaki*, provides clues to the changes in gene regulation that occurred during the transition to multicellularity.

DAVID S. BOOTH & NICOLE KING

The origin of all animals, from humans to sponges and comb jellies, can be traced back to a major event in evolutionary history: the transition to multicellularity. This transition was no doubt shaped by environmental changes — such as rising oxygen levels — and the evolution of cells that could engulf other, smaller cells<sup>1</sup>. However, to fully understand what drove this seminal event, we must look to the genome. Writing in *Cell*, Sebé-Pedrós *et al.*<sup>2</sup> report an investigation of gene regulation in a microscopic cousin of animals, *Capsaspora owczarzaki*. The study indicates that *Capsaspora* represents a transitional state in the evolution of gene-regulatory mechanisms, and provides a foundation for investigating how such mechanisms might have contributed to animal origins.

More than 600 million years ago, a series of genetic innovations allowed the progenitors of animals to exploit emerging environmental niches on a changing planet<sup>3</sup>. These progenitors cannot be studied directly, so how can we identify those genetic innovations that mattered most for animal origins?

Most insights into pre-animal genomes have come from comparisons of extant animals and their close relatives, choanoflagellates and *Capsaspora* (Fig. 1). Contrary to expectation, these studies revealed that much of the animal genetic toolkit (including the genes that encode cell-adhesion proteins such as integrins and cadherins, and those for vital signalling proteins such as receptor tyrosine kinases) is also expressed in *Capsaspora* and choanoflagellates<sup>4</sup>, indicating that many ‘animal’ genes pre-date animal origins.

Of course, animals are more than the sum of their genes — it is the regulated expression of genes across space and time that helps to differentiate egg from embryo, leg from wing or bat from fly. In plants and fungi, as well as in animals, transcription factors drive the synthesis of messenger RNA by interacting with regulatory regions called promoters that are located close to their target genes. Proximal control of transcription clearly pre-dates animal origins and is probably vital for all cellular life.

By contrast, long-range transcriptional regulation by DNA sequences called enhancers, which can lie more than 10 kilobases from the genes they regulate, has so far been

seen only in animals. Such regulation has been hypothesized to underlie the spatial and temporal coordination of cell differentiation that defines animal development<sup>5</sup>. But whether long-range enhancers are truly restricted to animals has been unclear, because they are often embedded in intricate transcriptional networks and can be difficult to detect.

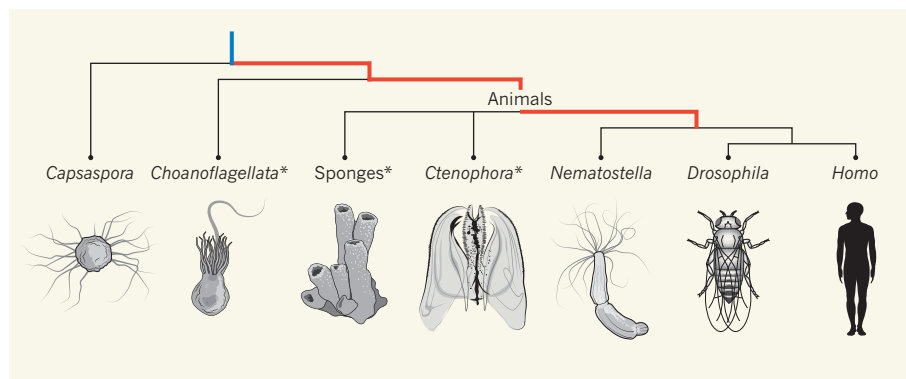
To investigate how different modes of transcriptional regulation may have set the stage for animal origins, Sebé-Pedrós *et al.* established approaches for functional genomics in *Capsaspora* (functional genomics probes how dynamic interactions between proteins, RNA and the genome correlate with gene expression). Despite the fact that *Capsaspora* is a non-model organism, it offers several benefits for such a study: it is easily cultured in the laboratory; it transitions between unicellular and aggregative multicellular forms; and its genome encodes many transcription factors that are evolutionarily conserved in animals<sup>6</sup>.

The authors report that, despite its relative simplicity, *Capsaspora* expresses two transcription factors that are integral to animal development — Myc and Brachyury. In animals, Myc serves as a master regulator of cell proliferation. Brachyury controls a key developmental process called gastrulation: this produces the body's three major cell layers, and the protein subsequently mediates differentiation of one of these layers, the mesoderm. In animals, both Myc and Brachyury function by binding to enhancers to regulate the transcription of a network of downstream genes<sup>7,8</sup>. Remarkably, Sebé-Pedrós *et al.* found that these downstream gene networks are conserved in animals and *Capsaspora*.

Given that cell proliferation is a shared feature of *Capsaspora* and animals, the conservation of the Myc regulatory network in the two lineages may not be surprising. But it is surprising that Brachyury seems to regulate the same types of gene in animals and *Capsaspora*, despite the fact that *Capsaspora* neither gastrulates nor produces mesoderm. Just as genes that animals use for cell adhesion and signalling evolved in the progenitors of animals before being co-opted for different functions in a multicellular context, it now seems that some gene-regulatory networks pre-date animal origins and were recruited wholesale for the regulation of new developmental processes.

Co-option is not the whole story, however. Innovations at the level of genes (such as that encoding the animal-specific signalling protein Wnt) and gene regulation (such as enhancer sequences) might also have contributed to animal origins. In contrast to the expansive intergenic DNA and long-range enhancers found in most animal genomes, the *Capsaspora* genome is compact. Despite looking for signatures of long-range transcriptional regulation at several stages of *Capsaspora*'s life cycle, Sebé-Pedrós *et al.* identified none.

Animals also seem to have evolved new



**Figure 1 | Evolution of gene-regulatory mechanisms.** Sebé-Pedrós *et al.*<sup>2</sup> report that two transcription factors, Myc and Brachyury, control similar sets of genes in animals and in a close relative, *Capsaspora owczarzaki*. This indicates that key gene-regulatory networks evolved before the origin of animals (indicated by the blue line) and were later co-opted for animal development. By contrast, long-range gene-regulatory elements called enhancers are not found in *Capsaspora*, but have been found in *Nematostella*, an animal that branched off early in evolutionary history. Thus, enhancers might be animal-specific (time window over which the evolution of long-range gene regulation might have occurred is indicated in red). A full understanding of how the animal gene-regulatory landscape evolved will require analyses of other early-branching animals such as sponges and *Ctenophora* (comb jellies), and other close relatives of animals, such as *Choanoflagellata*, in which gene regulation has not yet been studied (marked \*).

classes of promoter. Three types of animal promoter have been identified<sup>9</sup>: type I and type III promoters regulate genes that act during distinct stages in development, whereas type II promoters direct ubiquitous gene expression. Sebé-Pedrós and colleagues detected type II promoters in *Capsaspora*, but not types I or III. Therefore, type I and III promoters might be animal innovations.

It will be exciting to explore what these findings mean for animal origins and early evolution. Future investigations into the thus-far-uncharacterized gene-regulatory landscapes of sponges, comb jellies (ctenophores) and choanoflagellates promise to help pinpoint how and when long-range enhancers and type I and III promoters first evolved. However, the evolutionary distance between these organisms and the model animals that form the basis of our understanding of animal gene regulation may render conserved

molecular mechanisms unrecognizable by functional-genomic approaches. Moreover, other evolutionarily important gene-regulatory mechanisms may lie undiscovered in *Capsaspora*, choanoflagellates and animals that branched off early in the evolution of animals.

Fully reconstructing gene regulation in the progenitors of animals will require studies in diverse relatives, integrating modern functional genomics with forward and reverse genetics — which respectively reveal the genes responsible for a particular trait, and the changes brought about by disrupting the function of a particular gene. Fortunately, armed with the functional-genomics insights from this study, and the establishment of forward genetics in choanoflagellates<sup>10</sup>, this goal may be achieved in the not-too-distant future. ■

David S. Booth and Nicole King are at the Howard Hughes Medical Institute and in the

Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720-3200, USA. e-mails: dbooth@berkeley.edu; nking@berkeley.edu

1. Knoll, A. H. *Annu. Rev. Earth Planet. Sci.* **39**, 217–239 (2011).
2. Sebé-Pedrós, A. *et al. Cell* **165**, 1224–1237 (2016).
3. Erwin, D. H. *Biol. J. Linn. Soc.* **50**, 255–274 (1993).
4. Richter, D. J. & King, N. *Annu. Rev. Genet.* **47**, 509–537 (2013).
5. Levine, M., Cattoglio, C. & Tjian, R. *Cell* **157**, 13–25 (2014).
6. Sebé-Pedrós, A. *et al. eLife* **2**, e01287 (2013).
7. Lolas, M., Valenzuela, P. D. T., Tjian, R. & Liu, Z. *Proc. Natl Acad. Sci. USA* **111**, 4478–4483 (2014).
8. Hurlin, P. J. *Cold Spring Harb. Perspect. Med.* **3**, a014332 (2013).
9. Lenhard, B., Sandelin, A. & Carninci, P. *Nature Rev. Genet.* **13**, 233–245 (2012).
10. Levin, T. C., Greaney, A. J., Wetzel, L. & King, N. *eLife* **3**, e04070 (2014).

This article was published online on 15 June 2016.

## BIOGEOCHEMISTRY

# Synergy of a warm spring and dry summer

**An analysis suggests that high carbon uptake by US land ecosystems during the warm spring of 2012 offset the carbon loss that resulted from severe drought over the summer — and hints that the warm spring could have worsened the drought.**

YUDE PAN & DAVID SCHIMEL

Warmer springs and drier summers are an expected consequence of climate change<sup>1</sup>. Warmer springs should increase the carbon uptake of terrestrial ecosystems by lengthening the growing season, whereas drier summers should reduce uptake because of poor plant growth, especially in drought years. In 2012, the continental United States had the warmest spring on record, and one of the worst summer droughts in decades. What did these extremes do to the land carbon budget? The answer matters because terrestrial carbon uptake helps to remove anthropogenic carbon dioxide emissions from the atmosphere. Writing in *Proceedings of the National Academy of Sciences*, Wolf *et al.*<sup>2</sup> conclude that the increased carbon uptake during the spring essentially offset the carbon lost during the summer — although the details of this phenomenon are rather complex.

The effects of interactions between spring warming and summer drought on carbon budgets at continental and local scales have been reported previously<sup>3,4</sup>, but it is only in the past few years that multiple data sources with which to evaluate large-scale climate effects and their local variations have become

widely available. The authors arrived at their conclusions by comparing three data sets: eddy-covariance data that measure carbon exchange between the lowest part of the atmosphere (the boundary layer) and land biospheres over areas of approximately 1 square kilometre, gathered by 22 towers scattered across the United States; satellite estimates of the timing of plant growth; and regional carbon-budget estimates from CarbonTracker, a modelling system that uses observations of atmospheric CO<sub>2</sub> levels and gradients to infer surface fluxes of the gas over land. So what do the data show?

The severe drought that occurred during the summer of 2012 encompassed more than half of the continental United States, with most of the affected regions falling into the two worst categories as defined by the US Drought Monitor (extreme and exceptional)<sup>5</sup>. Accordingly, most of the towers reported a loss of carbon from their sites during this period, and recorded that the annual carbon budgets did not balance. Meanwhile, CarbonTracker suggested that carbon gain during the spring (0.24 petagrams of carbon; 1 Pg is 10<sup>15</sup> grams) and carbon loss during the summer (0.23 Pg) were almost equal for the continental United States as a whole.

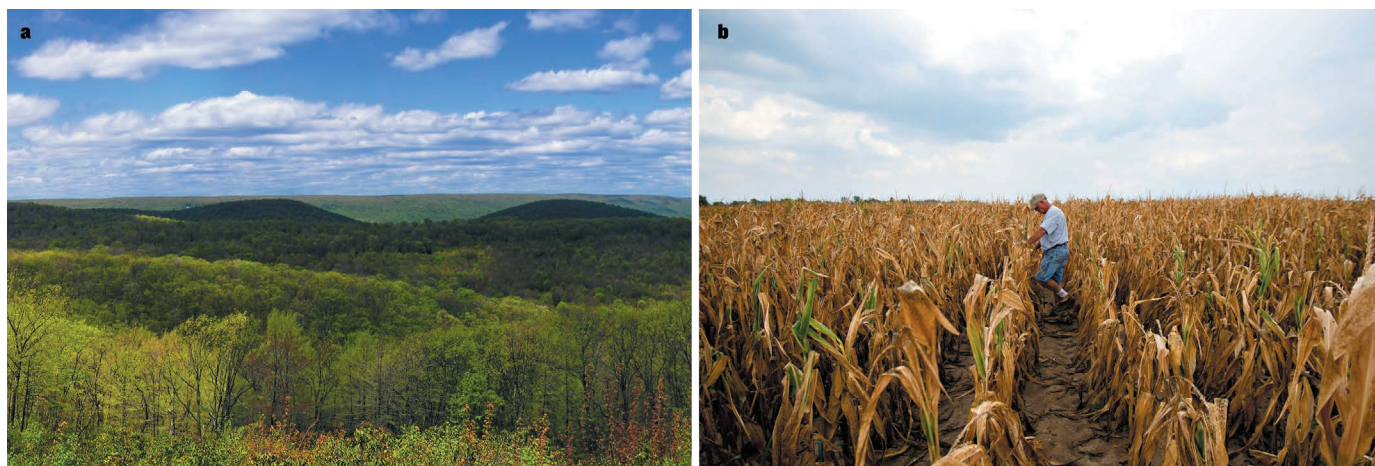
However, there was considerable variability

within that picture. Eastern temperate forests (Fig. 1a) vigorously sequestered carbon during the spring, and this carbon gain (0.18 Pg) slightly more than offset the summer carbon loss (0.16 Pg) from the Great Plains (Fig. 1b) — the area most affected by drought, and which accumulated significantly less carbon than in an average year. Overall, carbon uptake for the lands of the continental United States had increased, rather than reduced, by the end of the year (a rise of 0.11 Pg C yr<sup>-1</sup>), with the surplus resulting from increased carbon uptake during the autumn.

Wolf and colleagues propose that the spring warming and summer drought were physically coupled through interactions between the land surface and atmosphere. Simply put, ecosystems entered the summer with a relative water deficit because water was used up earlier than normal during the warmer spring. The deficit led to a reduction in evaporative cooling, which increased the effects of summer heating, causing water stress.

The authors go on to suggest that early warming might even have reinforced weather patterns, increasing the probability or the severity of summer drought. Confirming this will require a more comprehensive analysis and diagnosis, including measurements from more eddy-covariance towers, but is well within the realm of possibility. Clear evidence of such a link would undoubtedly help the public, policy-makers and resource managers to prepare strategies for adapting to droughts in the future.

A strength of Wolf and co-workers' study is that it combines *in situ* eddy-covariance measurements, atmospheric observations and remote-sensing data. The eddy-covariance data provide the most direct evidence for seasonal changes in terrestrial carbon uptake, and are the only data that directly constrain



**Figure 1 | Seasonal and regional variations of carbon uptake in the continental United States.** **a**, Eastern temperate forests grew vigorously during the warm spring of 2012, and took up more carbon than normal for this season. **b**, The subsequent hot, dry summer caused crops to fail in the Great Plains, and carbon uptake in this region was lower than normal. Wolf *et al.*<sup>2</sup> report that the spring carbon uptake offset the summer carbon losses across the continental United States.

the CarbonTracker and satellite estimates, by quantifying both the carbon flux and the full energy balance of water–temperature interactions. The remote-sensing data provide the best insight into the timing of biological activity across the continent, whereas the atmospheric analyses allow the local fluxes and processes to be understood in the context of the overall carbon budget. In the future, a more sophisticated synthesis of the different data will greatly improve the accuracy of analyses of carbon and water exchange between the land and atmosphere.

A limitation of the study is that the tower sites weren't specifically placed to sample the dominant carbon-flux anomalies that were revealed by CarbonTracker and the satellite data. For instance, the largest region of spring-time carbon-uptake anomalies occurred in the southeastern United States, where there are no flux towers. The largest region of midsummer carbon-loss anomalies occurred in the Great Plains, where the two sites used in the study represent grasslands, rather than the dominant agricultural landscapes of this region.

In addition, the current tower network isn't dense enough to cover climate events such as the extreme year of 2012. A facility called the National Ecological Observatory Network (with which one of us, D.S., was associated for several years), designed to sample climate conditions optimally, will come online in the next few years<sup>6</sup> and provide uniform coverage of the continental United States. Climatologists have long designed networks to study spatial patterns, whereas ecologists have tended to rely on local field studies and extrapolated their findings to larger areas on the basis of vegetation types or other classifications. A reference network that covers all spatial components and biomes is essential for this type of extrapolation in future studies.

Wolf and colleagues' work shows how important systematic, continental-scale sampling is, because no one site — and not even

several sites — could tell the entire story of a perturbation such as the one that occurred in 2012. As ecologists attempt to understand problems at ever larger scales, they will increasingly direct their creative energies towards problems that require massively more data than individual research laboratories can collect. Information obtained from infrastructural monitoring systems and openly available data will therefore have a crucial role in advancing the science of climate impacts, as they already do in other disciplines. ■

**Yude Pan** is in the Research Unit of Climate, Fire and Carbon Cycle Sciences, US Department of Agriculture Forest Service, Newtown Square, Pennsylvania 19073, USA.

**David Schimel** is at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA.  
e-mail: [ypan@fs.fed.us](mailto:ypan@fs.fed.us)

1. Melillo, J. M., Richmond, T. C. & Yohe, G. W. (eds) *Climate Change Impacts in the United States: The Third National Climate Assessment* <http://dx.doi.org/10.7930/J0Z31WJ2> (2014).
2. Wolf, S. *et al.* *Proc. Natl Acad. Sci. USA* **113**, 5880–5885 (2016).
3. Angert, A. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 10823–10827 (2005).
4. Hu, J., Moore, D. J. P., Burns, S. P. & Monson, R. K. *Glob. Change Biol.* **16**, 771–783 (2010).
5. Rippey, B. R. *Weather. Clim. Extremes* **10**, 57–64 (2015).
6. Keller, M., Schimel, D. S., Hargrove, W. W., Hoffman, F. M. *Front. Ecol. Environ.* **6**, 282–284 (2008).

This article was published online on 15 June 2016.

## STRUCTURAL BIOLOGY

# When sperm meets egg

**Sperm–egg binding is mediated by two cell–surface proteins. Structural analysis of these proteins, separately and in complex, provides insight into the recognition process and the subsequent sperm–egg fusion. SEE LETTERS P.562 & P.566**

**KARSTEN MELCHER**

**A**n interaction between two proteins — Izumo1, which is produced by sperm, and Juno, its receptor on eggs — enables human fertilization. However, the details of this interaction have been elusive. In two papers, Aydin *et al.*<sup>1</sup> (page 562) and Ohto *et al.*<sup>2</sup> (page 566) present the structures of Izumo1, Juno and the two proteins in complex, determined by X-ray crystallography at atomic-level resolution.

Following human copulation, motile sperm move towards eggs in the female's Fallopian tubes. The acidic environment of the female

reproductive tract triggers an activation step, in which sperm become hypermobile and penetrate the outer protective layer of the egg. A second activation step occurs when or shortly before the sperm binds to the zona pellucida — the tough inner layer that surrounds the egg. During this step, the acrosome — an organelle at the tip of the sperm head — releases digestive enzymes that break down the zona pellucida. This acrosome reaction allows the sperm to bind to Juno on the egg membrane, following which the two cells' membranes fuse and the cells merge. In turn, the egg releases enzymes that crosslink glycoproteins of the zona pellucida to make it impenetrable,

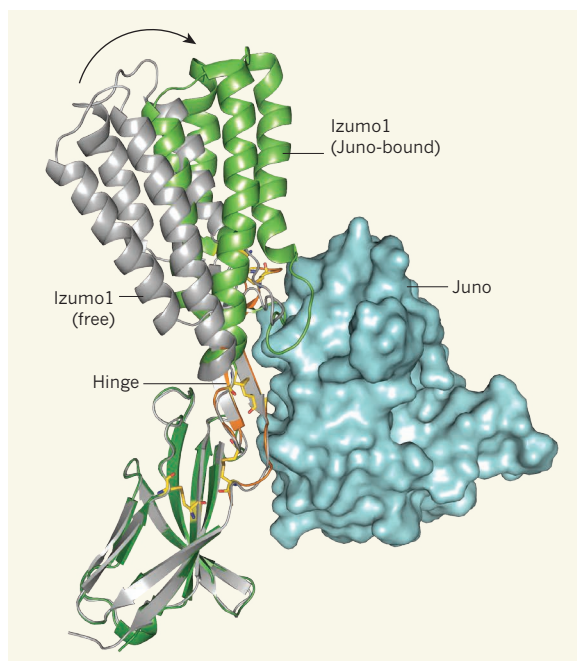
preventing fertilization by multiple sperm (polyspermy)<sup>3,4</sup>.

Izumo1, which is named after a Japanese marriage shrine, was first identified in 2005 by its binding to an antibody that blocked sperm–egg fusion<sup>5</sup>. The protein remains concealed intracellularly in the inner acrosomal membrane until the acrosome reaction occurs, when the inner membrane becomes part of the cell surface. Juno, named after the Roman goddess of love and marriage, was identified almost a decade later<sup>6</sup> as a membrane-anchored protein that is required for female fertility, sperm–egg membrane fusion, and egg binding by Izumo1. One structure of mouse Juno has been published this year<sup>7</sup>, and another will soon be published in *Nature Communications*<sup>8</sup>. But structures of the extracellular domain of Izumo1, human Juno and the Juno–Izumo1 complex have remained unknown.

Juno was originally called folate receptor- $\delta$ , and shares close to 60% amino-acid identity with human folate receptors<sup>6</sup> (receptors for folic acid and its derivatives). The structures of Juno from mice<sup>7,8</sup> and the current studies reveal that the protein has an almost identical fold to that of folate receptors<sup>9,10</sup>: globular, stabilized by eight disulfide bonds (S–S) and with a deep, ligand-binding pocket. But several key amino-acid residues in Juno's ligand-binding pocket differ from those of the folate receptors, consistent with the fact that Juno cannot bind folates<sup>4</sup>.

Both groups find that the extracellular region of Izumo1 has two domains — a four-helix bundle at the protein's amino terminus and an immunoglobulin-like domain at the carboxy terminus. The two domains are connected by a hinge region consisting of a  $\beta$ -hairpin structure with loops at either end that are anchored to the two folded domains by disulfide bonds. The researchers show that Izumo1 and Juno form a high-affinity complex in a 1:1 ratio. A surface of Juno distant from the pocket binds the outside of the hinge and makes contacts with both Izumo1 domains (Fig. 1).

Ohto and colleagues crystallized structures of free and Juno-bound Izumo1 in the same elongated conformation. By contrast, Aydin *et al.* report that Izumo1 alone adopts a boomerang-shaped conformation, in which the hinge is almost 40° more closed than that of Juno-bound Izumo1. The authors validated the approximate shape using a technique known as small-angle X-ray scattering. This provides low-resolution structural information about the protein in solution, thereby avoiding potential conformational biases that can arise in X-ray crystallography owing to



**Figure 1 | Juno stabilizes the Izumo1 hinge.** Aydin *et al.*<sup>1</sup> and Ohto *et al.*<sup>2</sup> have solved the structures of the human sperm protein Izumo1 and its egg receptor Juno. Izumo1 is shown in ribbon form and Juno in a surface representation. Izumo1 consists of two folded domains on either side of a connecting hinge (orange). When Izumo1 is in its free state, the hinge is more flexible and may allow the protein to adopt more-bent conformations than when it is bound to Juno (possible conformation change indicated by black arrow). Juno binding stabilizes the hinge, fixing it in an elongated conformation. This might expose disulfide bonds (S–S; yellow) for disulfide-exchange reactions to promote Izumo1 dimerization and subsequent sperm–egg membrane fusion.

crystal packing. These data indicate that the boomerang-shaped conformation is probably the predominant conformation of Izumo1 in solution. Moreover, although Juno binds to the outer hinge surface, the region most strongly stabilized by this binding seems to be inside the hinge. This suggests that the hinge can adopt different positions in Izumo1 alone, but that Juno fixes the conformation of Izumo1 by simultaneously binding to both domains.

Although binding interfaces are typically the most evolutionarily conserved surfaces of proteins, the Izumo1–Juno interface is less conserved than the remainder of either protein. Both groups suggest that variation at the binding surfaces might contribute to species specificity during fertilization, because sperm–egg fusions retain some specificity even if the zona pellucida (the main block to cross-species fertilization) is removed<sup>11</sup>. Ohto and colleagues introduced genetic mutations into mouse *Izumo1* that strongly reduced the affinity of the Izumo1–Juno interaction. Expression of wild-type *Izumo1* in monkey kidney cells (which do not normally express *Izumo1*) enabled these cells to bind efficiently to mouse eggs that lacked the zona pellucida, whereas cells that expressed the mutant protein could not. These results clearly confirm the interface

identified in these structures and its importance in mediating sperm–egg docking.

Why would a protein-binding receptor evolve from a folate receptor? It is tempting to speculate that an unidentified, non-folate ligand might bind the pocket of Juno to modulate the receptor's activity. Folate receptors are exquisitely pH-sensitive and release folic acid under acidic conditions<sup>10</sup>, and Ohto *et al.* demonstrated that slight acidification drastically decreased Juno's affinity for Izumo1. Together, ligand binding and pH changes could enable Juno to regulate Izumo1 binding at multiple levels.

Although the interaction between Izumo1 and Juno in sperm–egg recognition and adhesion has been structurally and biophysically characterized, the transition from initial binding to membrane fusion remains unclear. Izumo1 stays in the membrane following binding, whereas Juno is shed. This shedding might rapidly block polyspermy before the slow hardening of the zona pellucida is completed<sup>6</sup>. Previous work<sup>12</sup> suggests that Izumo1 undergoes stable dimerization through a disulfide-exchange reaction, dissociating from Juno to enable recruitment of membrane-fusion machinery. Indeed, Ohto *et al.* provide evidence that the disulfide bonds in Izumo1 are easily broken — perhaps stabilization of

Izumo1 following Juno binding could expose disulfides for exchange. Testing this hypothesis and determining how Izumo1–Juno binding triggers membrane fusion will require the identification of proteins that bind to Izumo1 after Juno shedding, and the reconstitution of events that follow initial binding in cells. ■

**Karsten Melcher** is at the Van Andel Research Institute, Grand Rapids, Michigan 49503, USA. e-mail: karsten.melcher@vai.org

1. Aydin, H., Sultana, A., Li, S., Thavalingam, A. & Lee, J. E. *Nature* **534**, 562–565 (2016).
2. Ohto, U. *et al.* *Nature* **534**, 566–569 (2016).
3. Krauchunas, A. R., Marcello, M. R. & Singson, A. *Mol. Reprod. Dev.* **83**, 376–386 (2016).
4. Okabe, M. *Development* **140**, 4471–4479 (2013).
5. Inoue, N., Ikawa, M., Isotani, A. & Okabe, M. *Nature* **434**, 234–238 (2005).
6. Bianchi, E., Doe, B., Goulding, D. & Wright, G. J. *Nature* **508**, 483–487 (2014).
7. Han, L. *et al.* *Curr. Biol.* **26**, R100–R101 (2016).
8. Kato, K. *et al.* *Nature Commun.* <http://dx.doi.org/10.1038/ncomms12198> (2016).
9. Chen, C. *et al.* *Nature* **500**, 486–489 (2013).
10. Wibowo, A. S. *et al.* *Proc. Natl Acad. Sci. USA* **110**, 15180–15188 (2013).
11. Yanagimachi, R., Yanagimachi, H. & Rogers, B. J. *Biol. Reprod.* **15**, 471–476 (1976).
12. Inoue, N., Hagihara, Y., Wright, D., Suzuki, T. & Wada, I. *Nature Commun.* **6**, 8858 (2015).

This article was published online on 15 June 2016.

# Searching for the rules that govern hadron construction

Matthew R. Shepherd<sup>1</sup>, Jozef J. Dudek<sup>2,3</sup> & Ryan E. Mitchell<sup>1</sup>

**Just as quantum electrodynamics describes how electrons are bound in atoms by the electromagnetic force, mediated by the exchange of photons, quantum chromodynamics (QCD) describes how quarks are bound inside hadrons by the strong force, mediated by the exchange of gluons. QCD seems to allow hadrons constructed from increasingly many quarks to exist, just as atoms with increasing numbers of electrons exist, yet such complex constructions seemed, until recently, not to be present in nature. Here we describe advances in the spectroscopy of mesons that are refining our understanding of the rules for predicting hadron structure from QCD.**

While decades of experimental study support QCD as the underlying theory of quark interactions, a detailed understanding of the way QCD generates protons, neutrons, and other strongly interacting ‘hadrons’ remains elusive. The majority of observed hadrons fall neatly into only two very limited sets: baryons, which are consistent with being three-quark constructions ( $qqq$ ); and mesons, which are quark–antiquark ( $q\bar{q}$ ) constructions. QCD also appears to allow constructions featuring larger numbers of quarks as well as hadrons built not only from quarks, but also from gluons. This has raised the question of why, until possibly now, there has been no evidence for a spectrum of such hadrons. Have we just been historically unsuccessful in producing these exotic particles in the laboratory, or are there more restrictive rules for building hadrons that are not obvious from the unsolved equations of QCD? Here we choose to focus specifically on the spectrum of mesons, where timely developments in both theory and experiment can be used to illustrate how the field of hadron spectroscopy addresses fundamental questions about QCD, questions that are common to both the meson and baryon sectors.

## Interacting quarks and gluons in QCD

Within QCD, the ‘charge’ that controls the interactions of quarks is known as ‘colour’, and it was the study of the empirical spectrum of hadrons that first introduced the concept of quarks and their threefold colour charge. Interactions in QCD are symmetric under changes of colour, that is, no single colour of quark behaves differently from the other two, and imposing this symmetry on the theory uniquely defines the interactions allowed in QCD between the quarks and the force-carrying gluons. Coloured quarks can interact by emitting or absorbing gluons, and because they carry colour charge themselves, gluons can also emit and absorb gluons.

Although observations about the spectrum of hadrons inspired the fundamental theory of quark interactions, calculating the detailed spectrum from this theory has so far been impossible. The difficulties in these calculations stem from the presence of gluon–gluon interactions, which make QCD forces very strong on the distance scale of  $10^{-15}$  m that characterizes hadrons. This ultimately results in a property called ‘confinement’, whereby quarks are permanently trapped inside composite hadrons, making it difficult to isolate the interaction of a single quark and antiquark from the collective behaviour of quarks and gluons in the hadron. The strong coupling means that, unlike for the electromagnetic force, where the exchange of two photons between electrons in an atom is far less probable than the exchange of just one, exchange of any number of

gluons between quarks in a hadron is every bit as probable as exchanging one. Because of this, there is no simple method of calculating the net effect of interactions between two quarks, and a QCD calculation of the mass of a hadron, easily measurable by experiment, becomes intractable.

## Understanding QCD via rules for building hadrons

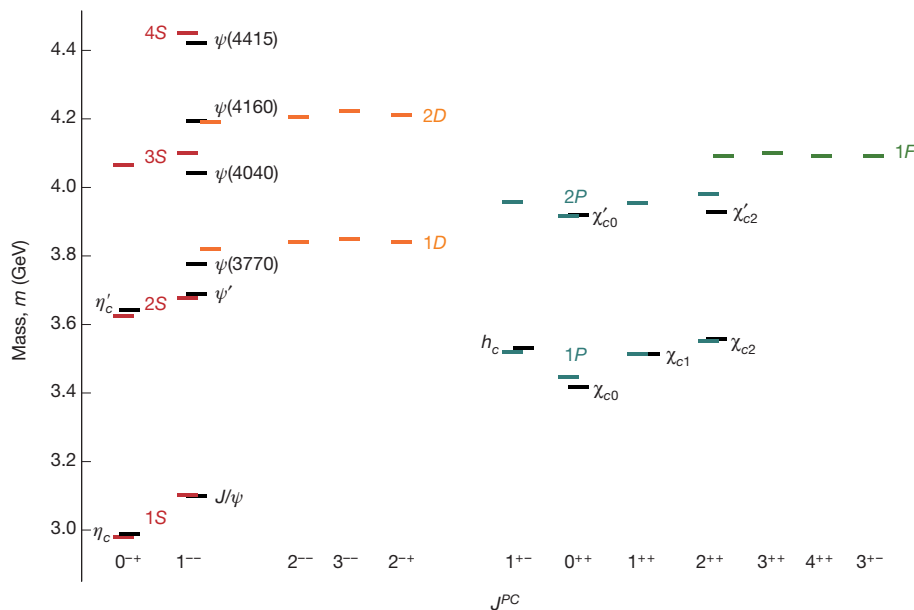
Our inability to solve the equations of QCD is not just a curiosity—it restricts our understanding of the behaviour and structure of hadrons, owing to the lack of any simple relationship between the fundamental quarks and gluons of QCD and the spectrum of hadrons observed experimentally. This has motivated the use of heuristic models, or ‘rules’, that serve as a bridge between QCD and experiment, capturing the important features of the spectrum while attempting to respect the known properties of QCD. The development of a rulebook for construction of hadrons consistent with both QCD and experimental data would arguably define what it means to understand how QCD generates hadrons. A uniform set of rules may not exist—there may be no simple way to capture the complex behaviour of QCD—but the high degree of regularity in the experimental spectrum of hadrons suggests that this is not a forlorn hope, and the search for this rulebook drives the field of hadron spectroscopy.

An important area of exploration attempts to create previously unobserved classes of hadrons in the laboratory, such as quark–gluon hybrids or tetraquarks. From the pattern of such states, or their absence, we can refine our understanding of the rules of hadron construction. A second area develops techniques for calculating the observable properties of hadrons directly from QCD, which will indicate how the rules follow from the strong interactions of quarks and gluons prescribed by that theory. In what follows we will review the current developments in each of these two areas and discuss the prospects for achieving the goal of determining the rulebook for hadron construction.

## Rules inferred from experimental data

We label hadrons by their mass and their quantum numbers  $J$  (spin),  $P$  (parity, behaviour under reflection in a mirror), and  $C$  (charge-conjugation, behaviour under exchange of particles with antiparticles). These properties are directly observable, but other characteristics, such as their internal composition, must be inferred. As the number of observed hadrons has increased over the last half-century, definite patterns have emerged that have led to an initial set of simple rules for the construction of hadrons from quarks.

<sup>1</sup>Department of Physics, Indiana University, Bloomington, Indiana 47405, USA. <sup>2</sup>Department of Physics, Old Dominion University, Norfolk, Virginia 23529, USA. <sup>3</sup>Jefferson Lab, Newport News, Virginia 23606, USA.



**Figure 1 | The charmonium spectrum.** A  $q\bar{q}$  potential model calculation (coloured)<sup>60</sup> of the charmonium spectrum is compared to experiment (black)<sup>61</sup>. Columns indicate states of common  $J^{PC}$ . Potential model states appear in groups labelled by their radial and orbital angular momentum quantum numbers,  $nL$  (coloured text;  $n = 1, 2, 3, \dots$ ;  $L = S, P, D, F, \dots$ ).

### The quark–antiquark rule for constructing mesons

One of the earliest patterns discovered (in the 1960s) was that mesons with the same  $J^{PC}$  quantum numbers could be grouped into sets of nine ('nonet') having similar mass. This could be explained by combining a quark  $q$  with an antiquark  $\bar{q}$  if there were three 'flavours' of quark—these were given the names 'up', 'down' and 'strange'. The lightest nonet of mesons has  $J^{PC} = 0^{-+}$ , and there are heavier nonets with other  $J^{PC}$  values. It was suggested that the additional mass-energy of the excited hadrons arises principally from the orbital or radial motion of the quark–antiquark ( $q\bar{q}$ ) pair, in analogy to the excitations of a single-electron atom.

With the discovery of charmonium (in the 1970s)<sup>1,2</sup>, this quantum-mechanical picture became more precise—these new mesons with masses much larger than those observed earlier were explained as being bound states featuring a new, heavier quark, which was dubbed 'charm'. Charmonium mesons with a range of  $J^{PC}$  values were observed and their spectrum (Fig. 1) resembles that of a pair of particles bound by a potential. The large mass of the charm quark justified such an approach, as many of the complexities of a relativistic system could be neglected. The potential needed to describe the spectrum was novel, featuring a steady rise at large distances that would confine the quarks within the meson<sup>3</sup>. A feature of this model of mesons is that it is not possible for a  $q\bar{q}$  pair in any orbitally or radially excited state to have  $J^{PC}$  in the set  $0^{+-}$ ,  $1^{-+}$ ,  $2^{+-}$ , .... Sets of mesons with these 'exotic' quantum numbers were not convincingly observed experimentally, either in charmonium or for the lighter quarks, supporting the  $q\bar{q}$  picture.

Until recently virtually all experimentally observed hadrons could have their presence explained by a simple rule stating that each meson is constructed from a  $q\bar{q}$  pair, and each baryon from a three-quark configuration. However, it has never been at all obvious why QCD is so parsimonious—why are there not meson-like states of two quarks and two antiquarks ('tetraquarks'), or baryon-like states of four quarks and an antiquark ('pentaquarks')? Furthermore, since the gluons of QCD strongly interact just as quarks do, could we not have 'hybrid mesons' in which gluons bind to a  $q\bar{q}$  pair, and 'glueballs' that do not require quarks at all? Observation of hadrons like these would challenge the simple rule outlined above, and indeed, recent experimental results are casting doubt on how parsimonious QCD really is.

### Recent results challenge the $q\bar{q}$ rule

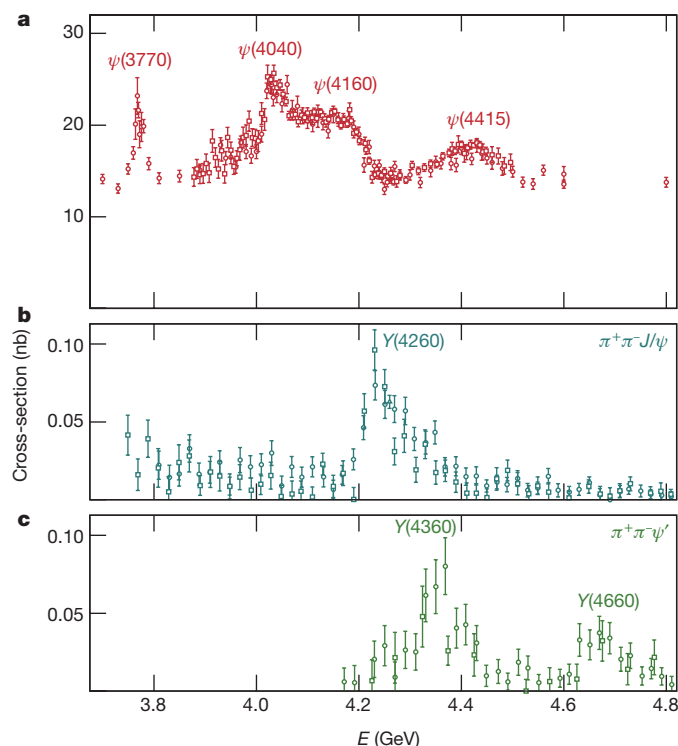
A powerful way to study the meson spectrum is to collide high-energy beams of electrons and positrons and to observe the rate at which systems

of hadrons are produced. In this process, the  $e^+e^-$  pair first annihilates, producing a photon; the photon converts into a quark and antiquark, which then interact, exchanging gluons and perhaps creating more  $q\bar{q}$  pairs; finally, these quarks and gluons arrange themselves into a system of hadrons that are observed by the particle detector. If the collision energy is close to the mass of a meson with  $J^{PC} = 1^{--}$  quantum numbers, the system 'resonates', and the probability of a collision increases. Thus, a plot of the normalized rate of hadron production, the 'cross-section', against the  $e^+e^-$  centre-of-mass energy, shows peaks corresponding to the produced meson states, also known as 'resonances' (Fig. 2a). These excited states exist only briefly before decaying into the set of observed lighter hadrons, and the width of the peak is inversely related to the lifetime of the state.

Figure 2a depicts the total rate of hadron production as a function of the  $e^+e^-$  centre-of-mass energy. The peaks are interpreted as evidence for a series of excited states—the  $\psi(3770)$ ,  $\psi(4040)$ ,  $\psi(4160)$  and  $\psi(4415)$ —consistent with expectations from the  $q\bar{q}$  picture (see Fig. 1). But recent experimental advances have allowed a closer inspection. If instead of the total rate, we look at the rates for the production of specific systems of hadrons, distinct features appear that have no simple explanation in the  $q\bar{q}$  picture.

The production rate of the  $\pi^+\pi^-J/\psi$  system, shown in Fig. 2b, provides one such example. (The  $J/\psi$  is a hadron that, for historical reasons, has two names associated with it,  $J$  and  $\psi$ .) Here, a prominent peak appears at 4,260 MeV, which, surprisingly, lies between the masses of the  $\psi(4160)$  and  $\psi(4415)$  states. Unlike the  $\psi(4160)$  and  $\psi(4415)$ , this  $Y(4260)$  resonance has no explanation within the  $q\bar{q}$  picture. Another example is the production rate of the  $\pi^+\pi^-\psi'$  system. The  $Y(4260)$  resonance might be expected also to appear here, since  $\pi^+\pi^-J/\psi$  and  $\pi^+\pi^-\psi'$  are very similar systems, but it does not. Instead, two peaks appear, for  $Y(4360)$  and  $Y(4660)$  (Fig. 2c), in further disagreement with the spectrum suggested by the total cross-section. These  $Y$  states, which appear in addition to those expected within the  $q\bar{q}$  picture, may be a signal that QCD does indeed produce mesons with internal structures beyond the simple  $q\bar{q}$  rule.

The observation of new states in charmonium, which was previously believed to be well understood, has spurred searches for further exotic candidate states, observations of which are providing still more challenges for the simple  $q\bar{q}$  rule. For example, a detailed study of the  $\pi^+\pi^-J/\psi$  system produced in  $Y(4260)$  decays showed that the  $\pi^\pm J/\psi$  system (Fig. 3a)



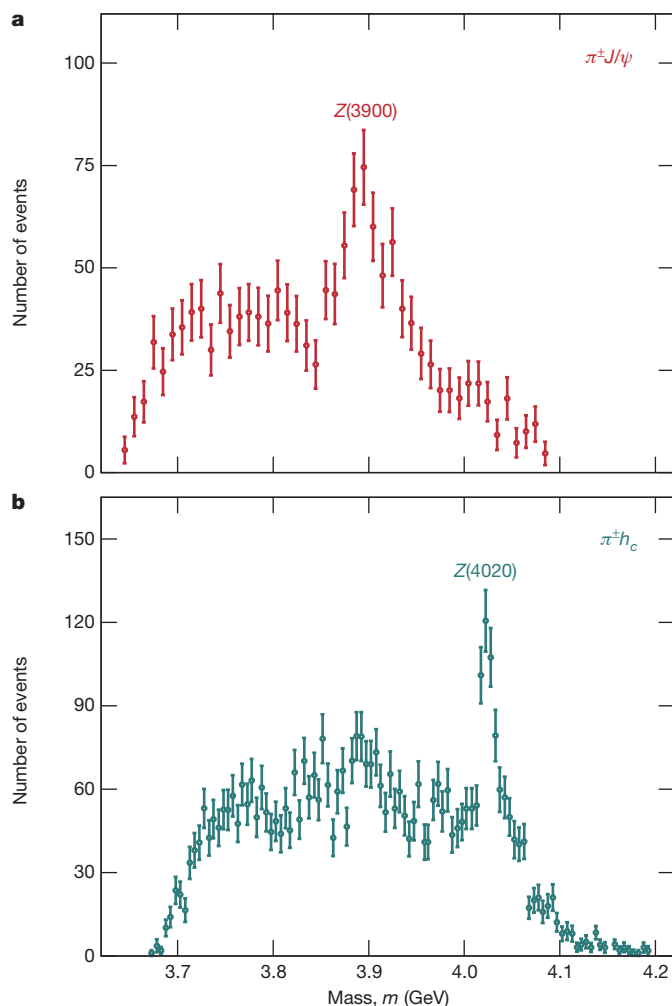
**Figure 2 | Electron-positron annihilation cross-sections.** **a**,  $e^+e^- \rightarrow$  hadrons (refs 62, 63). **b**,  $e^+e^- \rightarrow \pi^+\pi^- J/\psi$  (refs 4–6). **c**,  $e^+e^- \rightarrow \pi^+\pi^-\psi'$  (refs 65, 66). The  $1^{--}$  states  $\psi(3770)$ ,  $\psi(4040)$ ,  $\psi(4160)$  and  $\psi(4415)$ , indicated in **a**, can be associated with the  $1D$ ,  $3S$ ,  $2D$  and  $4S$  states of the potential model of Fig. 1. The error bars represent combined statistical and systematic uncertainties, taken from the appropriate references. The enhancements observed in **b** and **c** do not line up with these states, which may indicate that they correspond to new hadron states that do not appear in the potential model and hence do not obey the  $q\bar{q}$  rule. 1 nanobarn (nb) =  $10^{-37} \text{ m}^2$ .

appears to resonate at a mass of 3,900 MeV, producing an electrically charged state labelled  $Z(3900)^{4-6}$ . This structure is particularly noteworthy because its large mass and decay featuring  $J/\psi$  suggest that it contains a charm quark and an anti-charm quark, while its net electric charge requires further light (up- and down-flavoured) quarks. It is thus a possible tetraquark. A pattern of such states is beginning to emerge around 4 GeV: for example, in the  $\pi^+\pi^-h_c$  system, also produced in  $e^+e^-$  collisions, another electrically charged structure,  $Z(4020)$ , appears in the  $\pi^\pm h_c$  spectrum<sup>7</sup> (Fig. 3b) with a somewhat larger mass.

These new states can also, in principle, be produced in the weak decay of heavy mesons containing a bottom quark. Strangely, recent experimental data yields no evidence of  $Z(3900)$  production in such decays<sup>8</sup>. Instead, signals for still further new states of higher mass are observed<sup>8–11</sup>. A related process is the decay of heavy baryons containing a bottom quark, and here, equally as surprising, we find what appears to be a resonating proton- $J/\psi$  system<sup>12</sup>. This hadron is a possible pentaquark. Although the origin of these new states is not yet firmly established, they present a serious challenge to the simple rules for constructing mesons and baryons that we previously believed were obeyed by QCD.

The pattern of conventional mesons nicely replicates itself for each flavour of quark: many structures that appear in the spectrum of light quarks (up, down, strange) reappear for charm quarks at the 3-GeV scale, and again for bottom quarks at the 10-GeV scale. One might also expect that any spectrum of hybrids, tetraquarks or other novel constructions should have recurrent patterns for different quark flavours. In fact, bottom-quark analogues of the charged tetraquark candidates in charmonium have been reported<sup>13</sup>. Historically, these observations preceded those in charmonium.

Like tetraquarks and pentaquarks, another class of hadrons that appear to be allowed by the fundamental interactions of QCD are quark–gluon



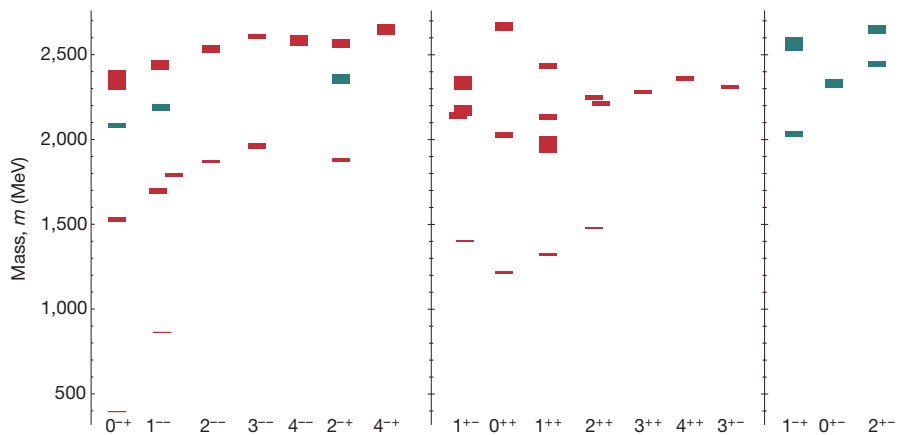
**Figure 3 | Number of events collected as a function of invariant mass.** **a**, For  $\pi^\pm J/\psi$  (ref. 5). **b**, For  $\pi^\pm h_c$  (ref. 7). In each case a clear narrow enhancement is observed that may be interpreted as a new hadron state. The error bars represent statistical uncertainty.

hybrids, in which gluons and quarks have a role in setting the quantum numbers of the hadron. A subset of possible hybrid mesons have a unique experimental signature: exotic  $J^{PC}$  not accessible to a  $q\bar{q}$  pair. While there are experimental indications of exotic hybrid candidates<sup>14–18</sup>, no firmly established spectrum of hybrid mesons has been discovered.

In parallel to the experimental work discussed above, theoretical efforts are underway to understand whether QCD predicts the existence of hadrons which go beyond the  $q\bar{q}$  meson and three-quark baryon rule, or whether the collective behaviour of quarks and gluons excludes the construction of more exotic combinations. It is to such calculations that our attention now turns.

## Rules derived from QCD

Much of our understanding of hadrons is informed by models, which may be motivated by features of QCD, by empirical observations, or both. A goal is to develop an understanding that is based on rigorous calculations of the interaction of quarks and gluons through the equations of QCD. However, the strongly coupled nature of QCD makes techniques that are practical for calculating weak and electromagnetic interactions ineffective for predicting properties of hadrons that emerge from QCD. We need a different approach, one that utilizes the fact that all fundamental particles, including quarks and gluons in QCD, are more correctly thought of as fluctuating quantum fields. The quantum aspect of the theory is embodied in the fact that observable consequences follow from a sum over all possible configurations in space and time that these fields can



**Figure 4 | Lattice QCD computation of the meson spectrum.** The spectrum is computed with light-quark masses such that  $m_\pi = 392$  MeV (ref. 67). The spectrum features sets of states compatible with the  $nL$  assignments of a  $q\bar{q}$  model (see Fig. 1), but also (shown in blue) states that do not have a place in such a model. These states can be interpreted as hybrid mesons in which a  $q\bar{q}$  pair is partnered with an excitation of the gluon field<sup>38</sup>—their presence suggests a new rule of hadron construction that includes gluons. (The height of each box represents the estimated uncertainty in the calculation.)

take. The method known as lattice QCD makes the approximation of considering these fields on a discrete grid of points describing a restricted region of space-time. A finite, but large, number of possible configurations of the fields on this grid can be generated using random sampling on a computer, and a good approximation for observable hadron properties obtained. The volume of the grid and number of field configurations required to achieve useful precision demands substantial computational resources. Total computational times of several teraflop-years are not unusual for contemporary calculations, with such efforts making use of ‘leadership-class’ supercomputing facilities—future precision lattice QCD calculations of increased sophistication will require petaflop-scale machines.

Lattice QCD has been applied with substantial success to a broad range of processes involving hadrons<sup>19</sup>, including the spectrum and internal structure of the lightest hadrons<sup>20</sup>, the behaviour of hadrons at non-zero temperature, relevant in collisions of heavy ions<sup>21</sup>, and heavy flavour decays, in which a heavy quark confined inside a meson decays through the weak interaction<sup>22</sup>.

### Lattice QCD as a tool for hadron spectroscopy

Our interest is in the determination of properties of excited hadrons, where obtaining a high degree of numerical precision is an issue that is secondary to the more basic question of whether certain states exist or do not. In the past few years we have seen excellent progress in overcoming the challenges posed by these calculations. Exploration of the excited hadron spectrum is possible using an approach in which each state in the spectrum is produced by a different combination of quark and gluon field constructions, and for this method to be successful, a large set of possible constructions is required. The dynamics of QCD, implemented by the sum over possible field configurations, determines which combination of constructions is present in each state in the spectrum. A scheme outlined in refs 23 and 24 includes many constructions resembling  $q\bar{q}$  pairs with various orbital motions and radial wavefunctions, motivated by the success of the  $q\bar{q}$  rule in describing the experimental hadron spectrum. More elaborate structures are possible, though, and refs 23 and 24 included several that feature the gluon field in a non-trivial way, inspired by the possibility that hybrid mesons may be allowed by QCD.

This large set of constructions, coupled with advances in computational techniques<sup>25</sup>, and the application of state-of-the-art computing hardware<sup>26,27</sup>, led to the pioneering results presented in Fig. 4 for the spectrum of mesons constructed from light up and down quarks. The computational challenges of these calculations currently require the utilization of masses for the lightest quarks that are heavier than the physical up and down quark masses, which leads to a systematic shift in the computed meson masses. However, since the immediate goal is to understand the underlying QCD dynamics by studying the pattern of states, rather than precisely to predict the mass of each meson, the computed spectrum allows us to develop intuitive rules for constructing hadrons that generally apply for quarks of any mass.

The spectrum presented in Fig. 4 qualitatively reproduces many of the features of the experimental light meson spectrum, and further it reflects the simple picture of  $q\bar{q}$  mesons, with the bulk of the states fitting into the pattern expected for states excited with increasing amounts of orbital angular momentum and/or excitations in the radial quantum number. There are some notable exceptions to this pattern, however, in particular the  $0^{-+}$ ,  $1^{-+}$  and  $2^{-+}$  states between 2.1 GeV and 2.4 GeV do not have an obvious explanation, and most strikingly there is a clear spectrum of states with exotic  $J^{PC} = 1^{-+}$ ,  $0^{+-}$  and  $2^{+-}$ , which cannot be constructed from a  $q\bar{q}$  pair alone.

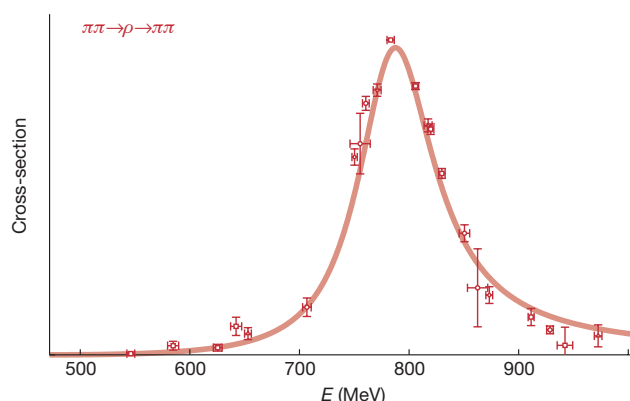
These additional mesons, which go beyond the set predicted by the  $q\bar{q}$  rule, have a natural explanation as quark–gluon hybrid mesons. Previously, estimates for the spectrum of hybrid mesons came only from models, which made educated guesses for the behaviour of the strongly coupled gluons inside a hadron. Different guesses led to very different predictions for the number and mass of hybrid states<sup>28–34</sup>. Using the lattice QCD technique, we are now able to predict a definitive pattern of states directly from the fundamental interactions as prescribed by QCD. Further calculations<sup>35–37</sup>, performed with larger values of the quark mass, up as high as the charm quark mass, show the same pattern of hybrid mesons, and they are found to be consistently 1.3 GeV heavier than the lightest  $J^{PC} = 1^{-+}$  meson. The particular pattern of states and the simple mass gap leads to a new rule of hadron construction for hybrid mesons, namely: combine  $q\bar{q}$  constructions with a gluonic field that has  $J^{PC} = 1^{+-}$  and a mass of about 1.3 GeV to form the spectrum of hybrid mesons in QCD. This is the first example of a rule following from a QCD calculation rather than being inferred from experimental observations<sup>38</sup>.

Of course this rule must be verified by producing and studying hybrid mesons in the laboratory, and many current and near-future experiments include searches for these states in their programmes. Some hybrid meson candidates have already been observed experimentally in both the light meson sector<sup>14–18</sup> and in the charm region. For example, the  $Y(4260)$  discussed in the previous section has  $J^{PC} = 1^{-+}$ , approximately the right mass relative to the  $J/\psi$ , and it seems to appear in addition to the expected  $q\bar{q}$  excitations. The new rule of hybrid meson construction would have this meson partnered with states of  $J^{PC} = (0, 1, 2)^{+-}$  at a similar mass. Searches for these states are underway.

### Calculating how hadrons decay

These calculations of the excited meson spectrum within QCD represent a major step forward in our understanding of hadron spectroscopy, but they still make approximations that fail to capture an important feature of excited hadrons—that they are resonances, decaying rapidly to lighter hadrons. As can be seen in Figs 2 and 3, in simple cases, excited states appear as characteristic peaks in the rate of observation of certain final-state mesons, and lattice QCD calculations should be capable of reproducing this behaviour.

Experimentally, resonances are often observed to decay preferentially into certain sets of mesons and not others, and these patterns can be used



**Figure 5 | Calculation of the  $\rho$  resonance.** The cross-section (in arbitrary units) is shown for  $\pi\pi \rightarrow \pi\pi$  with  $1^{--}$  quantum numbers, calculated using lattice QCD with light-quark masses such that  $m_\pi = 236$  MeV (ref. 68). The  $\rho$  resonance is clearly observed as a peak, and from its position and width the mass and decay rate of this excited state can be extracted. The errors represent the uncertainty in the calculation.

to infer details of the resonant state's internal structure. To be able to calculate the decay properties of the excited hadrons directly from QCD would provide a powerful tool for interpreting experimental data. In the case of predictions of previously unobserved excitations, it may also provide suggested decay channels to be examined in experimental searches.

Extending the calculations described above to account correctly for the decay of excited states is possible<sup>39–46</sup> but challenging, and serious efforts have only recently begun<sup>47–54</sup>. As an example of what can be achieved, in Fig. 5 we present the cross-section for two pions forming the lightest  $1^{--}$  resonance, known as the  $\rho$ , and then decaying back into two pions. A clear peak is observed, whose position and width provide the mass and decay rate of the  $\rho$ .

These rapidly maturing theoretical techniques will be required to study the new charmonium mesons, discussed earlier, within QCD. The observed enhancements are seen only in specific final states, which implies that the ability to predict how hadrons decay directly from QCD will be an essential component in interpreting experimental data in the quest to develop the rules for constructing hadrons.

### Towards a unified set of rules

Much of what we know about what emerges from strongly coupled QCD has come from studying patterns of hadrons organized by mass and quantum numbers like  $J$ ,  $P$  and  $C$ . These patterns suggest quarks of several flavours which may be combined with a single antiquark to form mesons—a rather simple rule of hadron construction. The theory of QCD is not limited to such simple constructions, however, and making a definitive statement about the existence of mesons with four-quark or

quark–gluon hybrid structure will require observing a spectrum of additional mesons that cannot be explained by the  $q\bar{q}$  rule. In particular, we will need to observe a set of states with unusual flavour and/or  $J^{PC}$  values.

### Finding a pattern of hadrons is essential

Contemporary technology has enabled experimental investigations at an unprecedented level of statistical precision, which provides the capability to discover more rare and interesting phenomena. However, we must exercise great care when we attempt to interpret experimental data. For example, one needs to be certain that the same logic that allows one to deduce the presence of conventional charmonium mesons in the total hadronic cross-section also applies when one is examining the cross-section for a single exclusive process that is two orders of magnitude smaller (see, for example, Fig. 2). Such precise experimental data make one susceptible to effects that can mimic the experimental signature of a new hadron, but which in fact may have a more prosaic origin<sup>55–59</sup>. This underscores the importance of experimentally establishing a pattern of hadrons: the interpretation of any single state as a new and exotic construction will certainly be questioned. However, the experimental observation of an ordered spectrum of states is harder to dismiss as a misinterpreted experimental artefact.

Likewise, theoretical efforts in lattice QCD must continue in their attempts to compute the complete set of possible hadrons allowed by QCD, and to identify patterns of states within that spectrum. Recent advances have enabled us to develop a simple rule, stated in subsection ‘Lattice QCD as a tool for hadron spectroscopy’, that describes how QCD constructs hybrid mesons and baryons, in an extension of what we had already for conventional mesons and baryons—this new rule must be verified by observing an experimental spectrum of hybrids. Lattice QCD can also be used to calculate decay properties of hadrons, and identifying particular characteristic decays of hybrid mesons will guide experimental searches and aid in interpretations of data. As has been done with hybrids, lattice QCD needs to determine whether QCD predicts a spectrum of tetraquark and pentaquark states. A particular priority is in the heavy quark sectors, where, as we have discussed above, there is recent experimental evidence for such objects. The ability within lattice QCD to vary arbitrarily the mass of the quarks allows us to identify how the rules of hadron construction vary, and to identify possible common behaviours between the heavy charmonium system and the lighter mesons.

### A global experimental programme

Establishing a spectrum of hadrons beyond those described by the simple  $q\bar{q}$  and  $qqq$  rules will require the combined efforts of multiple present and future experiments. There is a spectroscopy programme within nearly every particle physics collaboration worldwide. We list the details of a selection of several past, present, and future experiments, primarily those whose work is referenced in this article, in Box 1, as an illustration of the breadth of the worldwide effort.

#### BOX 1

### Hadron spectroscopy experiments

A selection of experiments and their hadron spectroscopy programmes, which typically represent only a fraction of each collaboration's research efforts.

**BaBar** (Menlo Park, California, USA):  $e^+e^-$  collisions at bottomonium energies; discoveries of the  $\Upsilon(4260)$  and  $\Upsilon(4360)$ ; finished collecting data in 2008.

**Belle** (Tsukuba, Japan):  $e^+e^-$  collisions at bottomonium energies; discovery of the  $X(3872)$ ,  $Z(3900)$ ,  $Z(4430)$ , and  $Z_b$  states; finished collecting data in 2010.

**Belle II** (Tsukuba, Japan): an upcoming continuation of the Belle experiment that will provide much higher intensity  $e^+e^-$  collisions than achieved at Belle.

**BESIII** (Beijing, China):  $e^+e^-$  collisions at charmonium energies; direct production of the  $\Upsilon(4260)$ ; discovery of the  $Z(3900)$  and  $Z(4020)$ ; ongoing.

**COMPASS** (Geneva, Switzerland): high-intensity meson beams on nuclear targets; searches for unusual light-quark mesons; discovery of the  $a_1(1420)$ ; ongoing.

**GlueX** (Newport News, Virginia, USA): polarized photon beam on a nuclear target; searches for light-quark hybrid mesons; data collection is beginning now.

**LHCb** (Geneva, Switzerland): high-energy, high-intensity proton–proton collisions, specializing in  $B$ -meson decays; measurement of resonant nature of the  $Z(4430)$ ; discovery of pentaquark candidates; ongoing.

**PANDA** (Darmstadt, Germany): proton–antiproton collisions at charmonium energies; exploration of charmonium and light-quark mesons; upcoming.

Most of the recently observed new hadrons have so far been observed in only a single production or decay process. Observation of the same state in multiple production and decay modes almost certainly rules out a misinterpretation of experimental data due to some process-dependent phenomenon and solidifies the evidence for a new hadron. Therefore the best current routes to explore new states in the charm sector are by comparing results from  $e^+e^-$  collisions (BESIII, Belle, Belle II) and production in  $B$ -meson decay (LHCb, Belle, Belle II). Supplementing these with results from novel production mechanisms, such as proton–antiproton annihilation ( $\bar{P}$ ANDA), would be extremely valuable.

Experiments aimed at exploring different energy regimes and quark flavours are essential for a complete understanding of the meson spectrum, as we expect the underlying patterns of states to be independent of quark mass. A variety of present and future experiments will allow access to both the charmonium system (BESIII, Belle, Belle II, LHCb), and the analogous system of bottom quarks, bottomonium (Belle, Belle II). Mesons constructed from light quarks can be produced in decays of heavier mesons and therefore can be studied at all of the previously mentioned facilities; they can also be produced at experiments dedicated to the study of lighter systems (COMPASS, GlueX). Discovery of light-quark hybrids would suggest the existence of heavy-quark hybrids and further motivate dedicated searches for these states.

With continued coordinated experimental and theoretical investigations we hope to define a complete set of rules for building hadrons that both describes what is observed in nature and can be derived directly from QCD. In doing so, we aim to understand how what seems to be a simple spectrum of hadrons emerges from the complex interactions of quarks and gluons in QCD.

Received 13 July 2015; accepted 31 March 2016.

1. Aubert, J. J. *et al.* Experimental observation of a heavy particle *J. Phys. Rev. Lett.* **33**, 1404 (1974).
2. Augustin, J. E. *et al.* Discovery of a narrow resonance in  $e^+e^-$  annihilation. *Phys. Rev. Lett.* **33**, 1406 (1974).
3. Eichten, E., Gottfried, K., Kinoshita, T., Lane, D. & Yan, T.-M. Charmonium: the model. *Phys. Rev. D* **17**, 3090 (1978).
4. Liu, Z. Q. *et al.* Study of  $e^+e^- \rightarrow \pi^+\pi^-J/\psi$  and observation of a charged charmoniumlike state at Belle. *Phys. Rev. Lett.* **110**, 252002 (2013).
5. Ablikim, M. *et al.* Observation of a charged charmoniumlike structure in  $e^+e^- \rightarrow \pi^+\pi^-J/\psi$  at  $\sqrt{s} = 4.26$  GeV. *Phys. Rev. Lett.* **110**, 252001 (2013). **This paper describes the observation of the Z(3900) at BESIII in decays to a charged pion and a  $J/\psi$ , simultaneously discovered at Belle<sup>4</sup>.**
6. Xiao, T., Dobbs, S., Tomaradze, A. & Seth, K. K. Observation of the charged hadron  $Z_c^\pm(3900)$  and evidence for the neutral  $Z_c^0(3900)$  in  $e^+e^- \rightarrow \pi\pi J/\psi$  at  $\sqrt{s} = 4170$  MeV. *Phys. Lett. B* **727**, 366 (2013).
7. Ablikim, M. *et al.* Observation of a charged charmoniumlike structure  $Z_c(4020)$  and search for the  $Z_c(3900)$  in  $e^+e^- \rightarrow \pi^+\pi^-h_c$ . *Phys. Rev. Lett.* **111**, 242001 (2013).
8. Chililkin, K. *et al.* Observation of a new charged charmoniumlike state in  $\bar{B}^0 \rightarrow J/\psi K^- \pi^+$  decays. *Phys. Rev. D* **90**, 112009 (2014).
9. Aaij, R. *et al.* Observation of the resonant character of the  $Z(4430)^-$  state. *Phys. Rev. Lett.* **112**, 222002 (2014).
10. Chililkin, K. *et al.* Experimental constraints on the spin and parity of the  $Z(4430)^+$ . *Phys. Rev. D* **88**, 074026 (2013).
11. Mizuk, R. *et al.* Observation of two resonance-like structures in the  $\pi^+\chi_{c1}$  mass distribution in exclusive  $\bar{B}^0 \rightarrow K^-\pi^+\chi_{c1}$  decays. *Phys. Rev. D* **78**, 072004 (2008).
12. Aaij, R. *et al.* Observation of  $J/\psi p$  resonances consistent with pentaquark states in  $\Lambda_b^0 \rightarrow J/\psi K^- p$  decays. *Phys. Rev. Lett.* **115**, 072001 (2015).
13. Bondar, A. *et al.* Observation of two charged bottomonium-like resonances in  $Y(5S)$  decays. *Phys. Rev. Lett.* **108**, 122001 (2012). **This paper describes the discovery of two  $Z_b$  peaks at Belle in decays to a charged pion and bottomonium.**
14. Meyer, C. A. & Swanson, E. S. Hybrid mesons. *Prog. Part. Nucl. Phys.* **82**, 21–58 (2015).
15. Ivanov, E. I. *et al.* Observation of exotic meson production in the reaction  $\pi^- p \rightarrow \eta^+ \pi^- p$  at 18 GeV/c. *Phys. Rev. Lett.* **86**, 3977 (2001).
16. Alekseev, A. *et al.* Observation of a  $J^{PC} = 1^{++}$  exotic resonance in diffractive dissociation of 190 GeV/c  $\pi^-$  into  $\pi^+ \pi^- \pi^+$ . *Phys. Rev. Lett.* **104**, 241803 (2010).
17. Kuhn, J. *et al.* Exotic meson production in the  $f_1(1285)\pi^-$  system observed in the reaction  $\pi^- p \rightarrow \eta^+ \pi^- \pi^- p$  at 18 GeV/c. *Phys. Lett. B* **595**, 109–117 (2004).
18. Adams, G. S. *et al.* Amplitude analyses of the decays  $\chi_{c1} \rightarrow \eta^+ \pi^- \pi^-$  and  $\chi_{c1} \rightarrow \eta^+ \pi^+ \pi^-$ . *Phys. Rev. D* **84**, 112009 (2011).
19. Kronfeld, A. S. Twenty-first century lattice gauge theory: results from the QCD Lagrangian. *Annu. Rev. Nucl. Part. Sci.* **62**, 265–284 (2012).
20. Kronfeld, A. S. in *100 Years of Subatomic Physics* (eds Henley, E. M. & Ellis, S. D.) 493–518 (World Scientific Publishing, 2013).
21. Ding, H.-T., Karsch, F. & Mukherjee, S. Thermodynamics of strong-interaction matter from lattice QCD. *Int. J. Mod. Phys. E* **24**, 1530007 (2015).
22. Aoki, S. *et al.* Review of lattice results concerning low-energy particle physics. *Eur. Phys. J. C* **74**, 2890 (2014).
23. Dudek, J. J., Edwards, R. G., Peardon, M. J., Richards, D. G. & Thomas, C. E. Highly excited and exotic meson spectrum from dynamical lattice QCD. *Phys. Rev. Lett.* **103**, 262001 (2009).
24. Dudek, J. J., Edwards, R. G., Peardon, M. J., Richards, D. G. & Thomas, C. E. Toward the excited meson spectrum of dynamical QCD. *Phys. Rev. D* **82**, 034508 (2010).
25. Peardon, M. *et al.* A novel quark-field creation operator construction for hadronic physics in lattice QCD. *Phys. Rev. D* **80**, 054506 (2009).
26. Egri, G. I. *et al.* Lattice QCD as a video game. *Comput. Phys. Commun.* **177**, 631–639 (2007).
27. Clark, M., Babich, R., Barrows, K., Brower, R. & Rebbi, C. Solving lattice QCD systems of equations using mixed precision solvers on GPUs. *Comput. Phys. Commun.* **181**, 1517 (2010).
28. Horn, D. & Mandula, J. A model of mesons with constituent gluons. *Phys. Rev. D* **17**, 898 (1978).
29. Isgur, N. & Paton, J. E. A flux tube model for hadrons in QCD. *Phys. Rev. D* **31**, 2910 (1985).
30. Jaffe, R. L., Johnson, K. & Ryzak, Z. Qualitative features of the glueball spectrum. *Ann. Phys.* **168**, 344–367 (1986).
31. Barnes, T., Close, F. E., de Viron, F. & Weyers, J. Q anti-Q hermaphrodite mesons in the MIT bag model. *Nucl. Phys. B* **224**, 241–264 (1983).
32. Chanowitz, M. S. & Sharpe, S. R. Hybrids: mixed states of quarks and gluons. *Nucl. Phys. B* **222**, 211 (1983); erratum *Nucl. Phys. B* **228**, 588 (1983).
33. General, I. J., Contanch, S. R. & Llanes-Estrada, F. J. QCD Coulomb gauge approach to hybrid mesons. *Eur. Phys. J. C* **51**, 347–358 (2007).
34. Guo, P., Szczepaniak, A. P., Galata, G., Vassallo, A. & Santopinto, E. Heavy quarkonium hybrids from Coulomb gauge QCD. *Phys. Rev. D* **78**, 056003 (2008).
35. Liu, L. *et al.* Excited and exotic charmonium spectroscopy from lattice QCD. *J. High Energy Phys.* **07**, 126 (2012).
36. Moir, G., Peardon, M., Ryan, S. M., Thomas, C. E. & Liu, L. Excited spectroscopy of charmed mesons from lattice QCD. *J. High Energy Phys.* **05**, 021 (2013).
37. Dudek, J. J. *et al.* Isoscalar meson spectroscopy from lattice QCD. *Phys. Rev. D* **83**, 111502 (2011).
38. Dudek, J. J. The lightest hybrid meson supermultiplet in QCD. *Phys. Rev. D* **84**, 074023 (2011). **This paper describes the interpretation of lattice QCD computed spectra, leading to a simple rule for the construction of hybrid mesons.**
39. Lüscher, M. Two particle states on a torus and their relation to the scattering matrix. *Nucl. Phys. B* **354**, 531 (1991).
40. Rummukainen, K. & Gottlieb, S. A. Resonance scattering phase shifts on a nonrest frame lattice. *Nucl. Phys. B* **450**, 397 (1995).
41. Kim, C. h., Sachrajda, C. T. & Sharpe, S. R. Finite-volume effects for two-hadron states in moving frames. *Nucl. Phys. B* **727**, 218 (2005).
42. Christ, N. H., Kim, C. & Yamazaki, T. Finite volume corrections to the two-particle decay of states with non-zero momentum. *Phys. Rev. D* **72**, 114506 (2005).
43. He, S., Feng, X. & Liu, C. Two particle states and the S-matrix elements in multi-channel scattering. *J. High Energy Phys.* **07**, 011 (2005).
44. Hansen, M. T. & Sharpe, S. R. Multiple-channel generalization of Lellouch-Lüscher formula. *Phys. Rev. D* **86**, 016007 (2012).
45. Briceño, R. A. & Davoudi, Z. Moving multichannel systems in a finite volume with application to proton-proton fusion. *Phys. Rev. D* **88**, 094507 (2013).
46. Guo, P., Dudek, J., Edwards, R. & Szczepaniak, A. P. Coupled-channel scattering on a torus. *Phys. Rev. D* **88**, 014501 (2013).
47. Aoki, S. *et al.* Lattice QCD calculation of the rho meson decay width. *Phys. Rev. D* **76**, 094506 (2007).
48. Feng, X., Jansen, K. & Renner, D. B. Resonance Parameters of the rho-meson from lattice QCD. *Phys. Rev. D* **83**, 094505 (2011).
49. Lang, C. B., Mohler, D., Prelovsek, S. & Vidmar, M. Coupled channel analysis of the rho meson decay in lattice QCD. *Phys. Rev. D* **84**, 054503 (2011); erratum *Phys. Rev. D* **89**, 059903 (2014).
50. Aoki, S. *et al.* rho meson decay in 2+1 flavor lattice QCD. *Phys. Rev. D* **84**, 094505 (2011).
51. Pelissier, C. & Alexandru, A. Resonance parameters of the rho-meson from asymmetrical lattices. *Phys. Rev. D* **87**, 014503 (2013).
52. Fu, Z. Lattice study on  $\pi K$  scattering with moving wall source. *Phys. Rev. D* **85**, 074501 (2012).
53. Lang, C. B., Leskovec, L., Mohler, D. & Prelovsek, S.  $K\pi$  scattering for isospin 1/2 and 3/2 in lattice QCD. *Phys. Rev. D* **86**, 054508 (2012).
54. Dudek, J. J., Edwards, R. G. & Thomas, C. E. Energy dependence of the rho resonance in  $\pi\pi$  elastic scattering from lattice QCD. *Phys. Rev. D* **87**, 034505 (2013); erratum *Phys. Rev. D* **90**, 099902 (2014).
55. Bugg, D. V. How resonances can synchronise with thresholds. *J. Phys. G* **35**, 075005 (2008).
56. Guo, F.-K., Meissner, U.-G., Wang, W. & Yang, Z. How to reveal the exotic nature of the  $P_c(4450)$ . *Phys. Rev. D* **92**, 071502 (2015).
57. Swanson, E. S.  $Z_b$  and  $Z_c$  exotic states as coupled channel cusps. *Phys. Rev. D* **91**, 034009 (2015).

58. Szczepaniak, A. P. Triangle singularities and XYZ quarkonium peaks. *Phys. Lett. B* **747**, 410–416 (2015).
59. Wang, Q., Hanhart, C. & Zhao, Q. Systematic study of the singularity mechanism in heavy quarkonium decays. *Phys. Lett. B* **725**, 106–110 (2013).
60. Godfrey, S. & Isgur, N. Mesons in a relativized quark model with chromodynamics. *Phys. Rev. D* **32**, 189 (1985).
61. Olive, K. A. *et al.* Review of particle physics. *Chin. Phys. C* **38**, 090001 (2014).
62. Bai, J. Z. *et al.* Measurements of the cross-section for  $e^+e^- \rightarrow$  hadrons at center-of-mass energies from 2 GeV to 5 GeV. *Phys. Rev. Lett.* **88**, 101802 (2002).
63. Osterheld, A. *et al.* Measurements of total hadronic and inclusive  $D^*$  cross-sections in  $e^+e^-$  annihilations between 3.87 GeV and 4.5 GeV. Stanford Linear Accelerator Center Report No. SLAC-PUB-4160, <http://slac.stanford.edu/pubs/slacpubs/4000/slac-pub-4160.pdf> (1986).
64. Lees, J. P. *et al.* (BaBar Collaboration) Study of the reaction  $e^+e^- \rightarrow J/\psi \pi^+ \pi^-$  via initial-state radiation at BaBar. *Phys. Rev. D* **86**, 051102 (2012).  
**This paper describes the discovery of the  $\Upsilon(4260)$  by the BaBar Collaboration in the process  $e^+e^- \rightarrow \pi^+ \pi^- J/\psi$ .**
65. Lees, J. P. *et al.* Study of the reaction  $e^+e^- \rightarrow \psi(2S) \pi^+ \pi^-$  via initial-state radiation at BaBar. *Phys. Rev. D* **89**, 111103 (2014).
66. Wang, X. L. *et al.* Observation of two resonant structures in  $e^+e^- \rightarrow \pi^+ \pi^- \psi(2S)$  via initial state radiation at Belle. *Phys. Rev. Lett.* **99**, 142002 (2007).
67. Dudek, J. J., Edwards, R. G., Guo, P. & Thomas, C. E. Toward the excited isoscalar meson spectrum from lattice QCD. *Phys. Rev. D* **88**, 094505 (2013).  
**This paper describes the detailed excited state spectra of light-quark mesons computed in lattice QCD.**
68. Wilson, D. J., Briceño, R. A., Dudek, J. J., Edwards, R. G. & Thomas, C. E. Coupled  $\pi\pi, K\bar{K}$  scattering in  $P$ -wave and the  $\rho$  resonance from lattice QCD. *Phys. Rev. D* **92**, 094502 (2015).  
**This paper describes the mapping out of the  $\pi\pi$  scattering amplitude and observation of the  $\rho$  resonance in lattice QCD.**

**Acknowledgements** J.J.D. acknowledges support provided by US Department of Energy contract DE-AC05-06OR23177, under which Jefferson Science Associates manages Jefferson Laboratory and the Early Career Award contract DE-SC0006765. M.R.S. and R.E.M. are supported by US Department of Energy contract DE-FG02-05ER41374. M.R.S. acknowledges the Jefferson Science Associates Sabbatical Leave Support Program. We thank L. Weinstein for comments on the initial draft of this manuscript.

**Author Contributions** All authors contributed equally to this manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.S. ([mashephe@indiana.edu](mailto:mashephe@indiana.edu)).

# Selective spider toxins reveal a role for the Na<sub>v</sub>1.1 channel in mechanical pain

Jeremiah D. Osteen<sup>1</sup>, Volker Herzig<sup>2</sup>, John Gilchrist<sup>3</sup>, Joshua J. Emrick<sup>1</sup>, Chuchu Zhang<sup>1</sup>, Xidao Wang<sup>4</sup>, Joel Castro<sup>5,6</sup>, Sonia Garcia-Caraballo<sup>5,6</sup>, Luke Grundy<sup>5,6</sup>, Grigori Y. Rychkov<sup>6</sup>, Andy D. Weyer<sup>7</sup>, Zoltan Dekan<sup>2</sup>, Eivind A. B. Undheim<sup>2</sup>, Paul Alewood<sup>2</sup>, Cheryl L. Stucky<sup>7</sup>, Stuart M. Brierley<sup>5,6</sup>, Allan I. Basbaum<sup>4</sup>, Frank Bosmans<sup>3</sup>, Glenn F. King<sup>2</sup> & David Julius<sup>1</sup>

**Voltage-gated sodium (Na<sub>v</sub>) channels initiate action potentials in most neurons, including primary afferent nerve fibres of the pain pathway. Local anaesthetics block pain through non-specific actions at all Na<sub>v</sub> channels, but the discovery of selective modulators would facilitate the analysis of individual subtypes of these channels and their contributions to chemical, mechanical, or thermal pain. Here we identify and characterize spider (*Heteroscodra maculata*) toxins that selectively activate the Na<sub>v</sub>1.1 subtype, the role of which in nociception and pain has not been elucidated. We use these probes to show that Na<sub>v</sub>1.1-expressing fibres are modality-specific nociceptors: their activation elicits robust pain behaviours without neurogenic inflammation and produces profound hypersensitivity to mechanical, but not thermal, stimuli. In the gut, high-threshold mechanosensitive fibres also express Na<sub>v</sub>1.1 and show enhanced toxin sensitivity in a mouse model of irritable bowel syndrome. Together, these findings establish an unexpected role for Na<sub>v</sub>1.1 channels in regulating the excitability of sensory nerve fibres that mediate mechanical pain.**

Pain is a multimodal system in which the activation of functionally distinct sensory nerve fibres elicits acute, protective reflexes as well as maladaptive responses that contribute to persistent pain<sup>1</sup>. In these nociceptive neurons, three voltage-gated sodium (Na<sub>v</sub>) channels—Na<sub>v</sub>1.7, Na<sub>v</sub>1.8 and Na<sub>v</sub>1.9—have garnered particular attention because mutations affecting these subtypes are associated with insensitivity to pain or persistent pain syndromes<sup>2–6</sup>. Na<sub>v</sub>1.1 is also expressed by somatosensory neurons<sup>7–10</sup>, but no link has been established between this subtype and nociception<sup>11</sup>. Mutations affecting Na<sub>v</sub>1.1 are associated with CNS disorders such as epilepsy<sup>12,13</sup>, autism<sup>14</sup>, and Alzheimer disease<sup>15</sup>, and these clinically dominant phenotypes may have masked functions of this subtype in peripheral neurons. For example, gain-of-function mutations in Na<sub>v</sub>1.1 underlie familial hemiplegic migraine type 3 (ref. 16), and although this phenotype has been ascribed to a CNS-initiated mechanism<sup>17</sup>, dysfunction in sensory neurons may also contribute to this pain syndrome.

Another challenge for identifying roles for Na<sub>v</sub>1.1 in pain is developing subtype-selective drugs for any member of this highly conserved family of ion channels<sup>18</sup>. Natural products can be exploited as a source of evolutionarily honed agents that target receptors with exquisite specificity. Such agents may be found in complex venoms from spiders, scorpions, cone snails, and snakes; they include toxins that excite sensory nociceptors to elicit pain or discomfort in offending predators<sup>19,20</sup>. Here we describe two algogenic tarantula toxins that selectively activate Na<sub>v</sub>1.1 to elicit acute pain and mechanical allodynia, providing new insights into specific roles of this channel and Na<sub>v</sub>1.1-expressing sensory nerve fibres in nociception and pain hypersensitivity.

## Selective Na<sub>v</sub>1.1-activating toxins

To identify novel toxins that target nociceptors, we used calcium imaging to screen more than 100 spider, scorpion and centipede venoms

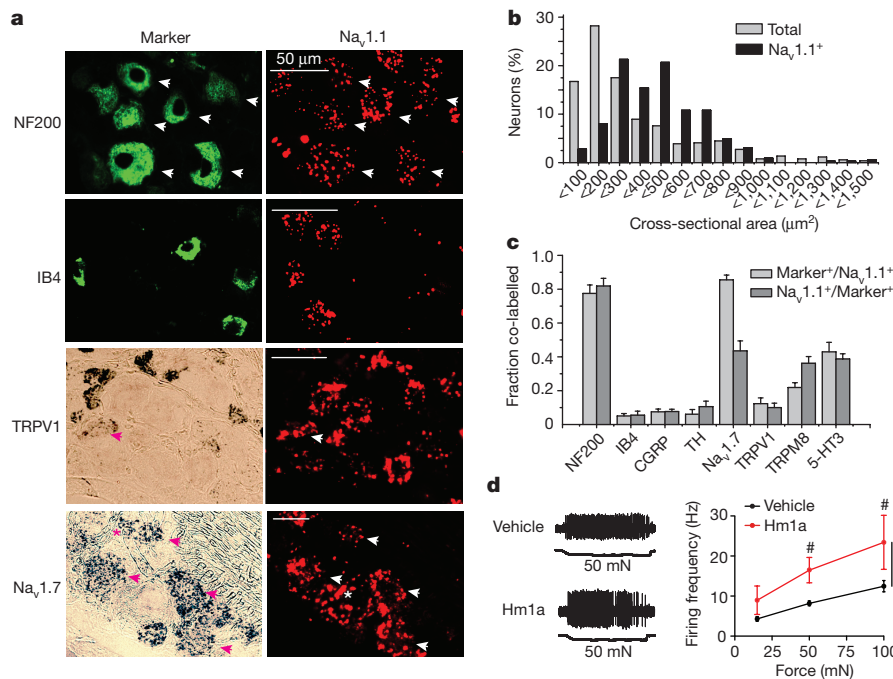
for the ability to activate cultured somatosensory neurons. Venom from the tarantula *Heteroscodra maculata* (Fig. 1a) robustly excited a subset of neurons from the trigeminal or dorsal root ganglia (DRG) of mice or rats. Venom fractionation yielded two active peptides, which were identified by matrix-assisted laser desorption/ionization–time-of-flight mass spectroscopy (MALDI–TOF MS) and Edman sequencing as inhibitor cystine knot (ICK) peptides with related sequences (Extended Data Fig. 1a). We named these toxins δ-theraphotoxin-Hm1a (Hm1a) and δ-theraphotoxin-Hm1b (Hm1b). Application of synthetic Hm1a to rat DRG neurons likewise triggered calcium responses (Fig. 1b), validating Hm1a as an active venom component. All subsequent experiments were performed with synthetic Hm1a peptide unless otherwise stated.

Tetrodotoxin (TTX) blocked Hm1a-evoked calcium responses (Fig. 1b), suggesting that these responses involved Na<sub>v</sub> channels. Indeed, whole-cell patch-clamp recordings from trigeminal neurons showed that Hm1a robustly inhibited Na<sub>v</sub> current inactivation (Fig. 1c). Among the Na<sub>v</sub> subtypes expressed by these neurons, only Na<sub>v</sub>1.1, 1.6 and 1.7 are sensitive to TTX<sup>21</sup>, narrowing our search. We next tested ICA-121431, a small molecule inhibitor with selectivity for the Na<sub>v</sub>1.1 and Na<sub>v</sub>1.3 subtypes<sup>22</sup> (Extended Data Fig. 1b), and found that it greatly diminished Hm1a-evoked calcium responses in both embryonic DRG and postnatal day (P)0 mouse trigeminal cultures (Fig. 1d and Extended Data Fig. 1c, d), suggesting that Na<sub>v</sub>1.1 is the main target of Hm1a in somatosensory neurons. In contrast, ICA-121431 only partially blocked responses to SGTx1, an Hm1a-related peptide that shows little selectivity among Na<sub>v</sub> subtypes<sup>23</sup> and excites a larger cohort of sensory neurons compared to Hm1a (Extended Data Fig. 1c, d). To confirm that the toxin was selective for Na<sub>v</sub>1.1 channels, we heterologously expressed Na<sub>v</sub>1.1–Na<sub>v</sub>1.8 α subunits in *Xenopus* oocytes. Hm1a potently inhibited inactivation of human (h)Na<sub>v</sub>1.1 channels (half-maximum effective concentration (EC<sub>50</sub>) = 38 ± 6 nM), but had

<sup>1</sup>Department of Physiology, University of California, San Francisco, California 94143, USA. <sup>2</sup>Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland 4072, Australia.

<sup>3</sup>Department of Physiology and Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. <sup>4</sup>Department of Anatomy, University of California, San Francisco, California 94143, USA. <sup>5</sup>Visceral Pain Group, Flinders University, Bedford Park, Southern Australia, 5042, Australia. <sup>6</sup>Centre for Nutrition and Gastrointestinal Diseases, Discipline of Medicine, University of Adelaide, South Australian Health and Medical Research Institute (SAHMRI), North Terrace, Adelaide, Southern Australia 5000, Australia.

<sup>7</sup>Department of Cell Biology, Neurobiology, and Anatomy, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA.



**Figure 3 | Na<sub>v</sub>1.1 is expressed by myelinated, non-C-fibre neurons in sensory ganglia.** **a**, Representative DRG sections showing immunoreactivity for neurofilament 200 (NF200), binding of isolectin B4 (IB4), and *in situ* histochemistry for TRPV1, Na<sub>v</sub>1.7, or Na<sub>v</sub>1.1 transcripts, as indicated. Arrows and asterisk indicate cells with overlapping and non-overlapping signals, respectively. **b**, Size distribution for all DRG neurons (grey bars, 514 cells counted) or Na<sub>v</sub>1.1-expressing cells (black bars, 324 cells counted). **c**, Quantification of overlap between histological markers

channel, which is normally insensitive to the toxin (Extended Data Fig. 2a). Transfer of just the DIV S3b–S4 region rendered rK<sub>v</sub>2.1 sensitive to Hm1a, demonstrating that this segment is a primary determinant of toxin action (Fig. 2b and Extended Data Fig. 2b). However, this region is identical or highly conserved in hNa<sub>v</sub>1.1, hNa<sub>v</sub>1.2 and hNa<sub>v</sub>1.3, and thus cannot fully account for toxin selectivity. To identify other functionally important regions, we constructed chimaeras between Na<sub>v</sub>1.1 and Na<sub>v</sub>1.4, which is completely insensitive to Hm1a. Replacement of the S3b–S4 region of rNa<sub>v</sub>1.4 with that of hNa<sub>v</sub>1.1 did not confer toxin sensitivity, whereas transfer of both S3b–S4 and the S1–S2 loop resulted in full toxin sensitivity (Fig. 2c and Extended Data Fig. 2c–e). These results indicate that both domains together determine toxin sensitivity and subtype selectivity, consistent with previous suggestions that S1–S2 contributes to toxin recognition sites on voltage sensors<sup>27,28</sup>.

### Na<sub>v</sub>1.1 is found on myelinated Aδ fibres

Using *in situ* hybridization histochemistry, we found that Na<sub>v</sub>1.1 transcripts were expressed primarily by medium-diameter sensory neurons (constituting 35% of all neurons within the DRG), most of which (>75%) belong to the myelinated (NF200-positive) cohort (Fig. 3). In contrast, we observed limited (5–11%) overlap of Na<sub>v</sub>1.1-positive cells with markers of small diameter, unmyelinated neurons, including the transient receptor potential cation channel subfamily V member 1 (TRPV1), calcitonin gene-related peptide (CGRP), tyrosine hydroxylase, and the lectin IB4. However, we did see substantial co-expression with the 5-HT<sub>3</sub> receptor, a marker of lightly myelinated Aδ neurons<sup>29</sup> (43% of Na<sub>v</sub>1.1-positive cells expressed 5-HT<sub>3</sub>). Finally, 22% of Na<sub>v</sub>1.1-positive cells also expressed the cold/menthol receptor TRPM8, which is found in both C and Aδ fibres<sup>30</sup>. From these findings, we conclude that Na<sub>v</sub>1.1 is expressed primarily by myelinated neurons, including Aδ fibres, consistent with previous histological and transcriptome profiling data<sup>7,31</sup>. We also characterized Hm1a-sensitive neurons for

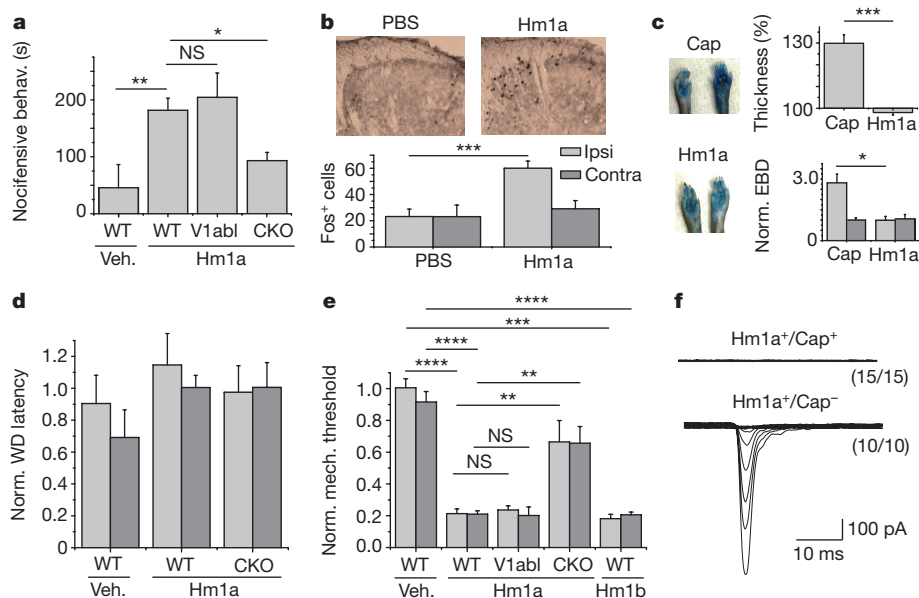
(≥164 cells counted for each condition; 9–12 independent sections from ≥3 mice). **d**, Representative traces from mechanonociceptive Aδ fibres recorded in skin-nerve preparation show increased firing following application of Hm1a (1 μM) with quantification on the right. Hm1a markedly increased firing during all forces tested, achieving statistical significance at 50 and 100 mN (\*\*\*)

responses to other receptor-selective agonists (Extended Data Fig. 3b), further confirming this conclusion. Notably, most (>85%) Na<sub>v</sub>1.1-positive cells also expressed Na<sub>v</sub>1.7, suggesting that this population of myelinated neurons contributes to nociception (see below).

We next investigated the effect of Hm1a on mechanonociceptive Aδ fibres using the *ex vivo* skin-nerve preparation. We found that application of 1 μM Hm1a to cutaneous receptive fields significantly increased the firing rate in mechanonociceptive Aδ fibres in response to mechanical stimuli (Fig. 3d), confirming expression of functional Na<sub>v</sub>1.1 channels in this afferent population. Previous studies found limited expression of TRPV1 in mechanonociceptive Aδ fibres<sup>32</sup>, consistent with our finding that there was limited overlap between Na<sub>v</sub>1.1 and TRPV1 expression. Together, these functional data confirm our histological assignment of Na<sub>v</sub>1.1 expression to myelinated Aδ fibres, and further suggest that Na<sub>v</sub>1.1 participates in mechanical nociception.

### Hm1a elicits pain and mechanical hypersensitivity

We investigated whether activation of Na<sub>v</sub>1.1-expressing fibres produced pain behaviours. Injection of Hm1a (5 μM in 10 μl) into the mouse hind-paw elicited immediate and robust nocifensive responses (bouts of licking or biting of the injected paw) throughout the observation period (Fig. 4a). Toxin injection also significantly increased Fos immunoreactivity in dorsal horn neurons of the superficial lamina ipsilateral to the injection, signifying functional engagement of nociceptors and their central connections (Fig. 4b). To exclude the possibility that this response depends on the small population of fibres that co-express TRPV1 and Na<sub>v</sub>1.1, we ablated TRPV1-positive terminals by intrathecal (spinal) injection of capsaicin<sup>33</sup>; Hm1a-evoked nocifensive behaviour persisted in these mice (Fig. 4a). Notably, Hm1a did not produce swelling or plasma extravasation of the injected paw, a neurogenic inflammatory response readily provoked by activation of peptidergic C-fibre nociceptors that include most TRPV1-expressing



**Figure 4 | Hm1a elicits non-inflammatory pain and bilateral mechanical allodynia.** **a**, Comparison of licking or biting behaviour (nocifensive behav.) following intraplantar injection (10  $\mu$ l) of vehicle (veh., PBS) ( $n=6$ ) versus Hm1a (5  $\mu$ M) ( $n=10$ ,  $**P<0.01$ ). Behaviour was unaffected by ablation of TRPV1 fibres (V1abl,  $n=5$ ) but significantly reduced in peripherin-Cre  $\times$  floxed-Nav1.1 (CKO) mice ( $*P<0.05$ ,  $n=11$ ). **b**, Top, representative histological sections and quantification of c-Fos immunoreactivity in spinal cord dorsal horn following intraplantar vehicle or Hm1a (5  $\mu$ M) injection ( $n=27$  sections from three mice,  $***P<0.001$ ). **c**, Capsaicin- or Hm1a-injected paws (right) next to uninjected contralateral controls (left). Top right, relative thickness of injected versus uninjected paws. Bottom right, Evans blue dye (EBD) extravasation following capsaicin or Hm1a injection ( $*P<0.05$ ).

**d**, Latency of paw withdrawal (WD) from noxious heat stimulus measured after intraplantar injection of vehicle or Hm1a (500 nM). **e**, Normalized mechanical response thresholds measured in paws ipsilateral (light grey) or contralateral (dark grey) to vehicle or toxin (500 nM) injection ( $n=5$  for wild type (WT) Veh., V1abl Hm1a and WT Hm1b;  $n=7$  for WT Hm1a;  $n=9$  for CKO Hm1a;  $**P<0.01$ ,  $***P<0.001$ ,  $****P<0.0001$ ). **f**, Mechanically evoked currents were observed from all adult mouse DRG neurons exhibiting sensitivity to Hm1a but not capsaicin (bottom), and not from those sensitive to both (top) (stimulus range 1–9  $\mu$ m displacement). Kinetic properties of mechanically evoked currents in Hm1a responders were variable. Error bars represent mean  $\pm$  s.e.m.  $P$  values based on unpaired two-tailed Student's  $t$ -test (**b**, **c**) or one-way ANOVA with post hoc Tukey's test (**a**, **d**, **e**).

neurons (Fig. 4c). These results further suggest that Hm1a elicits pain by activating a non-peptidergic subset of myelinated sensory fibres.

Genetic or pharmacological elimination of TRPV1-expressing fibres greatly diminishes sensitivity to noxious heat, but does not perturb sensitivity to mechanical stimuli<sup>33–35</sup>. In light of the anatomical and physiological results described above, we tested whether Hm1a has differential effects on these modalities by monitoring responses to thermal and mechanical stimuli following intraplantar injection of toxin at a dose (500 nM in 10  $\mu$ l) insufficient to elicit acute behaviour. Intraplantar injection of Hm1a did not alter sensitivity to heat, but produced robust sensitization to mechanical stimulation that was not dependent on TRPV1-expressing fibres (Fig. 4d, e). Equivalent mechanical sensitization was also observed following injection of native Hm1b peptide (Fig. 4e). Consistent with these behavioural observations, we found that all Hm1a-responsive adult DRG neurons displayed mechanically activated currents, except for those neurons that were also sensitive to capsaicin (Fig. 4f).

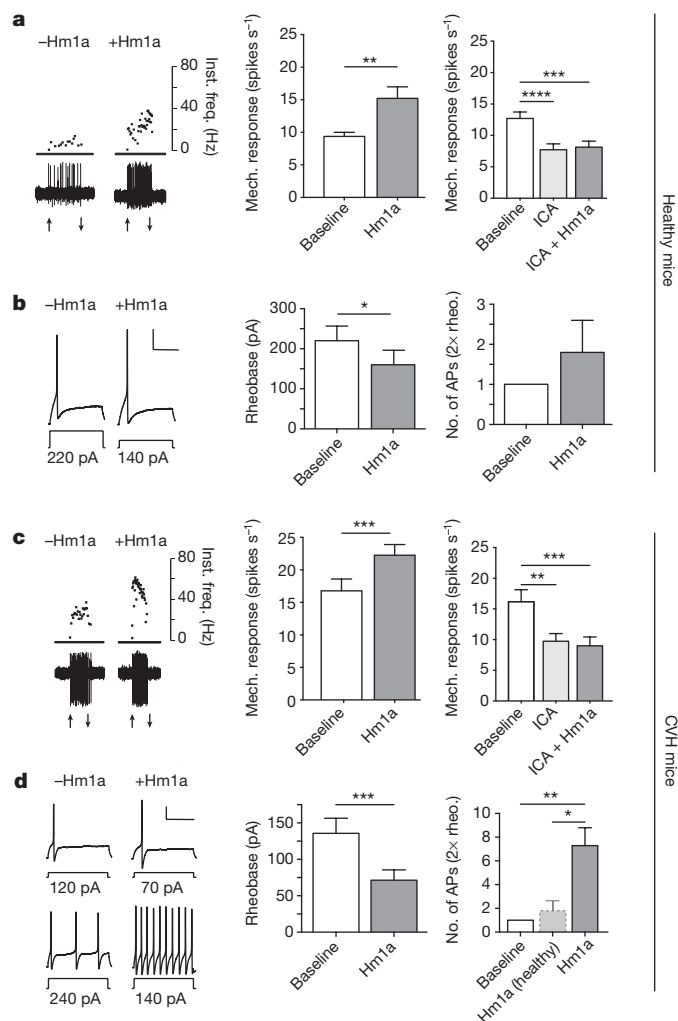
To confirm that Nav1.1 is required for toxin-evoked behaviours, we crossed mice bearing a floxed Nav1.1 allele<sup>13</sup> to a line expressing Cre recombinase under the control of the peripherin promoter, which is active in a large percentage of unmyelinated and myelinated sensory neurons during development<sup>36</sup>. Analysis of a peripherin-Cre  $\times$  yellow fluorescent protein (YFP) reporter line showed that these mice expressed Cre recombinase in 46% of Nav1.1-positive cells (Extended Data Fig. 4a, b). Notably, elimination of Nav1.1 from this subset of fibres significantly attenuated toxin-evoked behaviours, including both acute nocifensive responses and mechanical sensitization (Fig. 4a, e).

Robust activation of nociceptive pathways by nerve injury or inflammation can trigger both primary and secondary sensitization, the latter of which can manifest as mechanical or heat hypersensitivity contralateral to the insult<sup>37,38</sup>. In fact, we found that unilateral injection

of Hm1a produced robust and equivalent mechanical sensitization of both the injected and contralateral paws (Fig. 4e). This contralateral sensitization was also modality-specific, as no change in heat sensitivity was observed (Fig. 4d). Importantly, Hm1a-mediated mechanical sensitization was equivalently reduced in the ipsilateral and contralateral paws of Nav1.1-peripherin Cre mice, demonstrating that the contralateral effects depend on Nav1.1 (Fig. 4e). As we did not observe signs of neurogenic inflammation, we investigated whether this phenotype resulted from Hm1a-mediated nerve injury. However, this seems unlikely because toxin injection failed to induce expression of ATF3, a marker of nerve damage<sup>39</sup> (Extended Data Fig. 4c). Together, these observations demonstrate that direct activation of Nav1.1-expressing fibres is sufficient to produce robust and modality-specific bilateral sensitization.

### Nav1.1 and irritable bowel syndrome

Chronic mechanical hypersensitivity underlies the development of abdominal pain in patients with irritable bowel syndrome (IBS)<sup>40</sup>. Given the apparent role of Nav1.1 in mechanonociception, we investigated whether this channel is expressed by mechanically sensitive fibres of the gut, and, if so, whether it contributes to neuronal sensitization in a model of chronic visceral hypersensitivity (CVH)<sup>41</sup>. We examined mechanical responses in *ex vivo* gut-nerve preparations from healthy and CVH mice. In preparations from healthy mice, Hm1a increased mechanically evoked spiking in a subpopulation (40%) of high-threshold colonic afferents that constitute presumptive mechanonociceptors (Fig. 5a and Extended Data Fig. 5a). Correspondingly, ICA-121431 reduced mechanical responses in 50% of fibres examined and blocked Hm1a-induced sensitization (Fig. 5a and Extended Data Fig. 5a). Moreover, Hm1a significantly reduced the threshold for action potential firing in a subset (45%) of retrogradely



**Figure 5 | Colonic afferents display increased sensitivity to Hm1a in a model of IBS.** **a**, Left, representative *ex vivo* single fibre recording from an Hm1a (100 nM)-responsive high-threshold mechanoreceptive fibre from a healthy control mouse (arrows indicate application and removal of 2 g Von Frey hair stimulus). Middle, group data from Hm1a-sensitive fibres (6 out of 15, defined as  $\geq 15\%$  increase over baseline;  $**P < 0.01$ ; see Extended Data Fig. 5 for examples and group data from Hm1a-nonresponsive fibres). Right, group data from a population (5 out of 10) of ICA-121432 (500 nM)-sensitive afferents ( $****P < 0.0001$ ). **b**, Left, representative whole-cell current clamp recording of a retrogradely traced colonic DRG neuron in response to 500 ms current injection at rheobase (the minimum current injection required to elicit action potential firing). Recordings were from the same neuron of a healthy control mouse before and after incubation with Hm1a (10 nM). Scale bars, 250 ms, 20 mV. Middle, group data show significant reduction in rheobase following Hm1a application in a sub-population (5 out of 11) of neurons ( $*P < 0.05$ ). Hm1a-responsive neuron defined as exhibiting  $\geq 10\%$  change in rheobase from baseline control. Right, Hm1a increased the number of action potentials observed at  $2 \times$  rheobase in these neurons, but not to a level that reached statistical significance. **c**, Left, responses from high-threshold colonic fibres from CVH mice before and after application of Hm1a (100 nM). Middle, group data from Hm1a-responsive fibres (4 out of 11,  $***P < 0.001$ ). Right, group data from ICA-121432-sensitive fibres (7 out of 10,  $**P < 0.01$ ). **d**, Left, representative Hm1a-responsive colonic DRG neuron from CVH mice in whole-cell current clamp. Addition of Hm1a reduced rheobase (top traces) and increased action potential firing at  $2 \times$  rheobase (bottom traces). Middle and right, group data from Hm1a-responsive CVH neurons (7 out of 11) showing toxin-mediated decrease in rheobase (middle,  $***P < 0.001$ ) or increase in action potential firing at  $2 \times$  rheobase (right,  $*P < 0.05$ ,  $**P < 0.01$ ). Error bars represent mean  $\pm$  s.e.m. *P* values based on paired Student's *t*-test (**a–d**, middle) or one-way ANOVA with post hoc Bonferroni test (**a–d**, right).

traced colonic DRG neurons, as measured by whole-cell current clamp analysis (Fig. 5b and Extended Data Fig. 5b). These results demonstrate that a subset of high-threshold mechanosensitive colonic fibres express functional  $\text{Na}_v1.1$  channels.

In colonic afferents from CVH mice, baseline mechanosensory responses were elevated compared to healthy controls (compare Fig. 5a and c). Application of Hm1a enhanced mechanically evoked spiking in a subset (36%) of CVH fibres beyond this already elevated level (Fig. 5c). Notably, in the context of CVH (and in contrast to normal controls), toxin application markedly increased the electrical excitability of most (64%) retrogradely traced colonic DRG neurons (Fig. 5d), suggesting that  $\text{Na}_v1.1$  channels were functionally upregulated. Furthermore, ICA-121431 reduced mechanosensory responses in most (70%) CVH sensitized fibres to levels resembling those of baseline controls (compare Fig. 5a and c) and blocked the sensitizing effects of Hm1a (Fig. 5c). Together, these results support the idea that  $\text{Na}_v1.1$  is involved in mechanical hypersensitivity in IBS.

### Concluding remarks

The development of  $\text{Na}_v$  channel subtype-selective ligands is an important but challenging goal. Our results identify a site within the DIV S1–S2 loop that enhances subtype selectivity, providing a potential strategy for designing other subtype-specific gating modifiers. Moreover, toxins such as Hm1a and Hm1b, which alter inactivation, may be of particular utility in boosting  $\text{Na}_v$  channel activity where partial loss-of-function has been linked to developmental or neurodegenerative disorders, such as autism, Alzheimer disease and Dravet syndrome<sup>14,15,42</sup>. Analysis of toxin–channel interactions, including the multi-site nature of this pharmacophore, may shed new light on strategies for developing a broader class of molecules with similar selectivity and functional profiles.

The critical role of  $\text{Na}_v1.1$  in the brain may have prevented its prior recognition as a contributor to peripheral pain signalling. Our results now unambiguously implicate  $\text{Na}_v1.1$  and  $\text{Na}_v1.1$ -expressing myelinated afferents in nociception. Activation or sensitization of these fibres is sufficient to elicit robust acute pain and mechanical allodynia without triggering neurogenic inflammation, distinguishing these fibres from well-characterized C-nociceptors. Previous studies have implicated myelinated A $\delta$  fibres in mechanonociception<sup>43,44</sup>, and  $\text{Na}_v1.1$  now provides an important new marker with which to more precisely identify the contribution of these fibres to acute and chronic pain.

Our experiments in CVH mice suggest that pharmacological blockade of  $\text{Na}_v1.1$  represents a novel therapeutic strategy for diminishing the chronic pain in IBS, and perhaps other pain conditions associated with mechanical sensitization, including migraine headache. Although  $\text{Na}_v1.1$  activity in the brain may underlie aura in patients with type 3 familial hemiplegic migraine (FHM3)<sup>17</sup>, our results suggest that these gain-of-function mutations may also produce migraine pain through actions of  $\text{Na}_v1.1$  in mechanical nociceptors. In fact, anticonvulsants that target  $\text{Na}_v$  channels, including  $\text{Na}_v1.1$ , have been shown to reduce migraine attacks in some individuals<sup>45,46</sup>. Moreover, rufinamide, an anticonvulsant that was recently shown to inhibit  $\text{Na}_v1.1$  (ref. 47), has also been reported to diminish nerve-injury-evoked mechanical allodynia<sup>48</sup>. Our findings provide a mechanistic rationale for these actions, and motivate further analysis of the roles of  $\text{Na}_v1.1$  and  $\text{Na}_v1.1$ -expressing nociceptors in acute and persistent pain.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 October 2015; accepted 6 April 2016.

Published online 6 June 2016.

1. Basbaum, A. I., Bautista, D. M., Scherrer, G. & Julius, D. Cellular and molecular mechanisms of pain. *Cell* **139**, 267–284 (2009).

2. Dib-Hajj, S. D., Yang, Y., Black, J. A. & Waxman, S. G. The Na<sub>v</sub>1.7 sodium channel: from molecule to man. *Nat. Rev. Neurosci.* **14**, 49–62 (2013).
3. Faber, C. G. *et al.* Gain-of-function Na<sub>v</sub>1.8 mutations in painful neuropathy. *Proc. Natl Acad. Sci. USA* **109**, 19444–19449 (2012).
4. Fertleman, C. R. *et al.* SCN9A mutations in paroxysmal extreme pain disorder: allelic variants underlie distinct channel defects and phenotypes. *Neuron* **52**, 767–774 (2006).
5. Hoeijmakers, J. G., Faber, C. G., Merkies, I. S. & Waxman, S. G. Painful peripheral neuropathy and sodium channel mutations. *Neurosci. Lett.* **596**, 51–59 (2015).
6. Zhang, X. Y. *et al.* Gain-of-function mutations in SCN11A cause familial episodic pain. *Am. J. Hum. Genet.* **93**, 957–966 (2013).
7. Fukuoka, T. *et al.* Comparative study of the distribution of the  $\alpha$ -subunits of voltage-gated sodium channels in normal and axotomized rat dorsal root ganglion neurons. *J. Comp. Neurol.* **510**, 188–206 (2008).
8. Black, J. A., Liu, S., Tanaka, M., Cummins, T. R. & Waxman, S. G. Changes in the expression of tetrodotoxin-sensitive sodium channels within dorsal root ganglia neurons in inflammatory pain. *Pain* **108**, 237–247 (2004).
9. Fukuoka, T., Miyoshi, K. & Noguchi, K. De novo expression of Na<sub>v</sub>1.7 in injured putative proprioceptive afferents: Multiple tetrodotoxin-sensitive sodium channels are retained in the rat dorsal root after spinal nerve ligation. *Neuroscience* **284**, 693–706 (2015).
10. Dib-Hajj, S. D., Cummins, T. R., Black, J. A. & Waxman, S. G. Sodium channels in normal and pathological pain. *Annu. Rev. Neurosci.* **33**, 325–347 (2010).
11. Wang, W., Gu, J., Li, Y. Q. & Tao, Y. X. Are voltage-gated sodium channels on the dorsal root ganglion involved in the development of neuropathic pain? *Mol. Pain* **7**, 16 (2011).
12. Catterall, W. A., Kalume, F. & Oakley, J. C. Na<sub>v</sub>1.1 channels and epilepsy. *J. Physiol. (Lond.)* **588**, 1849–1859 (2010).
13. Cheah, C. S. *et al.* Specific deletion of Na<sub>v</sub>1.1 sodium channels in inhibitory interneurons causes seizures and premature death in a mouse model of Dravet syndrome. *Proc. Natl Acad. Sci. USA* **109**, 14646–14651 (2012).
14. Han, S. *et al.* Autistic-like behaviour in *Scn1a*<sup>+/−</sup> mice and rescue by enhanced GABA-mediated neurotransmission. *Nature* **489**, 385–390 (2012).
15. Verret, L. *et al.* Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model. *Cell* **149**, 708–721 (2012).
16. Gargus, J. J. & Tournay, A. Novel mutation confirms seizure locus SCN1A is also familial hemiplegic migraine locus FHM3. *Pediatr. Neurol.* **37**, 407–410 (2007).
17. Cestèle, S., Schiavon, E., Rusconi, R., Franceschetti, S. & Mantegazza, M. Nonfunctional Na<sub>v</sub>1.1 familial hemiplegic migraine mutant transformed into gain of function by partial rescue of folding defects. *Proc. Natl Acad. Sci. USA* **110**, 17546–17551 (2013).
18. Priest, B. T. Future potential and status of selective sodium channel blockers for the treatment of pain. *Curr. Opin. Drug Discov. Devel.* **12**, 682–692 (2009).
19. Bohlen, C. J. & Julius, D. Receptor-targeting mechanisms of pain-causing toxins: How ow? *Toxicol.* **60**, 254–264 (2012).
20. Olivera, B. M., Hillyard, D. R., Marsh, M. & Yoshikami, D. Combinatorial peptide libraries in drug design: lessons from venomous cone snails. *Trends Biotechnol.* **13**, 422–426 (1995).
21. Black, J. A. *et al.* Spinal sensory neurons express multiple sodium channel  $\alpha$ -subunit mRNAs. *Brain Res. Mol. Brain Res.* **43**, 117–131 (1996).
22. McCormack, K. *et al.* Voltage sensor interaction site for selective small molecule inhibitors of voltage-gated sodium channels. *Proc. Natl Acad. Sci. USA* **110**, E2724–E2732 (2013).
23. Bosmans, F., Martin-Eauclaire, M. F. & Swartz, K. J. Deconstructing voltage sensor function and pharmacology in sodium channels. *Nature* **456**, 202–208 (2008).
24. Bosmans, F., Puopolo, M., Martin-Eauclaire, M. F., Bean, B. P. & Swartz, K. J. Functional properties and toxin pharmacology of a dorsal root ganglion sodium channel viewed through its voltage sensors. *J. Gen. Physiol.* **138**, 59–72 (2011).
25. Escoubas, P., Diochot, S., Célérier, M. L., Nakajima, T. & Lazdunski, M. Novel tarantula toxins for subtypes of voltage-dependent potassium channels in the K<sub>v</sub>2 and K<sub>v</sub>4 subfamilies. *Mol. Pharmacol.* **62**, 48–57 (2002).
26. Smith, J. A. M., Davis, C. L. & Burgess, G. M. Prostaglandin E<sub>2</sub>-induced sensitization of bradykinin-evoked responses in rat dorsal root ganglion neurons is mediated by cAMP-dependent protein kinase A. *Eur. J. Neurosci.* **12**, 3250–3258 (2000).
27. Wang, J. *et al.* Mapping the receptor site for  $\alpha$ -scorpion toxins on a Na<sup>+</sup> channel voltage sensor. *Proc. Natl Acad. Sci. USA* **108**, 15426–15431 (2011).
28. Bende, N. S. *et al.* A distinct sodium channel voltage-sensor locus determines insect selectivity of the spider toxin Dc1a. *Nat. Commun.* **5**, 4350 (2014).
29. Zeitz, K. P. *et al.* The 5-HT<sub>3</sub> subtype of serotonin receptor contributes to nociceptive processing via a novel subset of myelinated and unmyelinated nociceptors. *J. Neurosci.* **22**, 1010–1019 (2002).
30. Bautista, D. M. *et al.* The menthol receptor TRPM8 is the principal detector of environmental cold. *Nature* **448**, 204–208 (2007).
31. Usoskin, D. *et al.* Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015).
32. Stucky, C. L., Medler, K. A. & Molliver, D. C. The P2Y agonist UTP activates cutaneous afferent fibers. *Pain* **109**, 36–44 (2004).
33. Cavanaugh, D. J. *et al.* Distinct subsets of unmyelinated primary sensory fibers mediate behavioral responses to noxious thermal and mechanical stimuli. *Proc. Natl Acad. Sci. USA* **106**, 9075–9080 (2009).
34. Mishra, S. K., Tisel, S. M., Orestes, P., Bhargava, S. K. & Hoon, M. A. TRPV1-lineage neurons are required for thermal sensation. *EMBO J.* **30**, 582–593 (2011).
35. Mishra, S. K. & Hoon, M. A. Ablation of TrpV1 neurons reveals their selective role in thermal pain sensation. *Mol. Cell. Neurosci.* **43**, 157–163 (2010).
36. Scherrer, G. *et al.* VGLUT2 expression in primary afferent neurons is essential for normal acute pain and injury-induced heat hypersensitivity. *Proc. Natl Acad. Sci. USA* **107**, 22296–22301 (2010).
37. Koltzenburg, M., Wall, P. D. & McMahon, S. B. Does the right side know what the left is doing? *Trends Neurosci.* **22**, 122–127 (1999).
38. Kissin, I., Lee, S. S. & Bradley, E. L. Jr. Effect of prolonged nerve block on inflammatory hyperalgesia in rats: prevention of late hyperalgesia. *Anesthesiology* **88**, 224–232 (1998).
39. Tsujino, H. *et al.* Activating transcription factor 3 (ATF3) induction by axotomy in sensory and motoneurons: A novel neuronal marker of nerve injury. *Mol. Cell. Neurosci.* **15**, 170–182 (2000).
40. Brierley, S. M. & Linden, D. R. Neuroplasticity and dysfunction after gastrointestinal inflammation. *Nat. Rev. Gastroenterol. Hepatol.* **11**, 611–627 (2014).
41. de Araujo, A. D. *et al.* Selenoether oxytocin analogues have analgesic properties in a mouse model of chronic abdominal pain. *Nat. Commun.* **5**, 3165 (2014).
42. Jensen, H. S., Grunnet, M. & Bastlund, J. F. Therapeutic potential of Na<sub>v</sub>1.1 activators. *Trends Pharmacol. Sci.* **35**, 113–118 (2014).
43. Khan, G. M., Chen, S. R. & Pan, H. L. Role of primary afferent nerves in allodynia caused by diabetic neuropathy in rats. *Neuroscience* **114**, 291–299 (2002).
44. Tsumozaki, M. *et al.* A ‘toothache tree’ alkylamide inhibits A $\delta$  mechanonociceptors to alleviate mechanical pain. *J. Physiol. (Lond.)* **591**, 3325–3340 (2013).
45. Pappagallo, M. Newer antiepileptic drugs: possible uses in the treatment of neuropathic pain and migraine. *Clin. Ther.* **25**, 2506–2538 (2003).
46. Calabresi, P., Galletti, F., Rossi, C., Sarchielli, P. & Cupini, L. M. Antiepileptic drugs in migraine: from clinical aspects to cellular mechanisms. *Trends Pharmacol. Sci.* **28**, 188–195 (2007).
47. Gilchrist, J. *et al.* Na<sub>v</sub>1.1 modulation by a novel triazole compound attenuates epileptic seizures in rodents. *ACS Chem. Biol.* **9**, 1204–1212 (2014).
48. Suter, M. R., Kirschmann, G., Laedermann, C. J., Abriel, H. & Decosterd, I. Rufinamide attenuates mechanical allodynia in a model of neuropathic pain in the mouse and stabilizes voltage-gated sodium channel inactivated state. *Anesthesiology* **118**, 160–172 (2013).
49. Herzig, V. *et al.* ArachnoServer 2.0, an updated online resource for spider toxin sequences and structures. *Nucleic Acids Res.* **39**, D653–D657 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the Deutsche Arachnologische Gesellschaft and particularly I. Wendt, J. Broghammer, A. Schlosser, B. Rast, M. Luescher, C. and F. Schneider and H. Auer for providing arthropods for milking; W. Catterall for providing floxed Na<sub>v</sub>1.1 mice; and J. Poblete, J. Madder, T. O'Donnell and A. Harrington for technical assistance. This work was supported by a T32 Postdoctoral Training Grant from the UCSF CVRI (J.D.O.), Ruth Kirschstein NIH postdoctoral (F32NS081907 to J.D.O.) and predoctoral (F31NS084646 to J.G. and F30DE023476 to J.J.E.) fellowships, the National Institutes of Health (R37NS065071 and R01NS081115 to D.J., R01NS091352 to F.B., R01NS040538 and R01NS070711 to C.L.S., and R37NS014627 and R01DA29204 to A.I.B.), the National Health and Medical Research Council of Australia (Project Grant APP1083480 to S.M.B., Program Grant APP1072113 and Principal Research Fellowship APP1044414 to G.F.K.), and a grant from the Wellcome Trust to A.I.B. S.M.B. is a NHMRC R.D. Wright Biomedical Research Fellow.

**Author Contributions** J.D.O., V.H., E.A.B.U., G.F.K. and D.J. carried out venom collection and screening, toxin purification and characterization. Z.D. and P.A. carried out Hm1a synthesis. J.D.O., J.G., C.Z., D.J. and F.B. designed, performed, and analysed electrophysiological and calcium imaging experiments to determine toxin mechanism and selectivity. J.J.E., J.D.O., X.W. A.I.B. and D.J. designed, performed and analysed histological experiments. A.D.W. and C.L.S. designed, performed, and analysed skin-nerve recordings. X.W., J.D.O., D.J., and A.I.B. designed, performed, and analysed behavioural experiments to assess somatic function. J.C., S.G.-C., L.G., G.Y.R. and S.M.B. designed, performed and analysed studies relating to colonic afferent and patch clamp pharmacological studies. All authors contributed to the discussion and interpretation of the results. J.D.O. and D.J. wrote the manuscript with contributions and suggestions from all authors.

**Author Information** Sequences and activity profiles for Hm1a and Hm1b can be found in the ArachnoServer database<sup>49</sup> under accession numbers AS000224 and AS002363, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.J. ([david.julius@ucsf.edu](mailto:david.julius@ucsf.edu)), G.K. ([glenn.king@imb.uq.edu.au](mailto:glenn.king@imb.uq.edu.au)) or F.B. ([frankbosmans@jhmi.edu](mailto:frankbosmans@jhmi.edu)).

## METHODS

**Venom collection and screening.** Venom from spiders, scorpions and centipedes was collected by mild electrical stimulation of the chelicerae, telson or forcipules, respectively. Venom samples were then lyophilized and kept frozen until use. Approximately 100 venoms were tested by ratiometric calcium imaging using a standard inverted microscope setup. Responses to high extracellular potassium (150 mM), capsaicin (1  $\mu$ M), or previously characterized venoms or purified toxins<sup>19</sup> were used to validate the health and robustness of sensory neuron cultures used in screening assays. Responses were digitized and analysed using MetaMorph software (Molecular Devices). Venom-evoked responses that were stimulus-locked, visually detectable above background, and restricted to neurons (that is, did not cause calcium entry into glia or fibroblasts) were selected for further analysis. Pharmacological analysis was used to narrow down potential targets and crude venoms or purified fractions were subsequently tested on cloned candidate channels. Candidates were taken forward based on the robustness of the response and evidence for selectivity at novel targets. See Supplementary Information for a summary of venoms that produced no detectable or specific response in our hands.

**Hm1a/b isolation.** Venom from *H. maculata* (1 mg dried) was fractionated on a C18 reversed-phase (RP) high-performance liquid chromatography (HPLC) column (Jupiter 250  $\times$  4.6 mm, 5 mm; Phenomenex) on a Shimadzu Prominence HPLC system. The following linear gradient of solvent B (90% acetonitrile (MeCN), 0.1% formic acid in water) in solvent A (0.1% formic acid in water) was used at a flow rate of 1 ml min<sup>-1</sup>: 5% B for 5 min, then 5–20% B for 5 min followed by 20–40% B over 40 min. Absorbance was measured at 214 nm and 280 nm and collected fractions were lyophilized before storage at –20°C.

**Mass spectrometry.** Peptide masses were determined by MALDI-TOF MS using a 4700 Proteomics Bioanalyzer model (Applied Biosystems). Peptides were dissolved in water and mixed 1:1 (v/v) with  $\alpha$ -cyano-4-hydroxycinnamic acid matrix (7 mg ml<sup>-1</sup> in 50% MeCN, 5% formic acid) and mass spectra acquired in positive reflector mode. All reported masses are for the monoisotopic M+H<sup>+</sup> ions.

**Sequence determination.** N-terminal sequencing was performed by the Australian Proteome Analysis Facility. In brief, Hm1a (600 pmol) and Hm1b (250 pmol) were reconstituted and reduced by addition of DTT (25 mM) followed by incubation at 56°C for 0.5 h. The samples were then alkylated using iodoacetamide (55 mM) at room temperature for 0.5 h and purified by RP-HPLC using a Zorbax 300SB-C18 column (3  $\times$  150 mm). The target peaks of interest were identified, collected and then reduced to minimal volume under vacuum. The entire sample was loaded onto a precycled, Biobrene-treated disc and subjected to 37 (Hm1a) or 42 (Hm1b) cycles of Edman N-terminal sequencing. Automated Edman degradation was carried out using an Applied Biosystems 494 Procise Protein Sequencing System.

Edman degradation of Hm1a yielded ECRYLFGGCSSTSDCKHLSCRSWDW KYCAWDGTF as the sequence, which has a calculated monoisotopic mass (for the M+H<sup>+</sup> ion) of 3,908.58 Da. This is 86.97 Da short of the monoisotopic mass of Hm1a of 3,995.55 Da. Hence, we concluded that a serine residue (87 Da) was missing from the C-terminal end of Hm1a to give a complete sequence of ECRYLFGGCSSTSDCKHLSCRSWDWKYCAWDGTF. The complete sequence has a calculated monoisotopic mass (for the M+H<sup>+</sup> ion) of 3,995.61 Da, which is only 0.06 Da different from the mass that was measured for the native Hm1a.

Edman degradation of Hm1b yielded ECRYLFGGCKTTADCKHLGCRTDLY YCAWDGT as the sequence, which has a calculated monoisotopic mass (for the M+H<sup>+</sup> ion) of 3,745.6 Da. This is 147 Da short of the monoisotopic mass of Hm1a of 3,892.60 Da. We therefore concluded that an amidated phenylalanine was missing from the C-terminal end of Hm1b to give a complete sequence of ECRYLFGGCKTTADCKHLGCRTDLYYCAWDGTF-NH<sub>2</sub>. To confirm that the C terminus of Hm1b is amidated, we digested native Hm1b with carboxypeptidase Y (CPY) and monitored the reaction by MALDI-TOF MS to identify the mass of the C-terminal residue as described previously<sup>50</sup>. Native Hm1b (5  $\mu$ l of 800 ng  $\mu$ l<sup>-1</sup>) in 100 mM ammonium acetate, pH 5.5, was incubated with 2 ng  $\mu$ l<sup>-1</sup> CPY at 37°C for 20 min. The reaction was monitored by removing 0.4  $\mu$ l at 0, 1, 5, 10 and 20 min and spotting it on a MALDI plate with equal volume of 7 mg ml<sup>-1</sup>  $\alpha$ -cyano-4-hydroxycinnamic acid in 60% (v/v) MeCN, 5% formic acid (FA). Dried spots were washed with 10  $\mu$ l 1% FA and allowed to dry before they were analysed by MALDI-TOF MS on a 4700 Proteomics Bioanalyzer (Applied Biosciences), acquiring spectra in reflector positive mode. The first CPY-mediated cleavage yielded a mass difference of 146 Da, which corresponds to an amidated phenylalanine residue. Thus, the complete sequence has a calculated monoisotopic mass (for the M+H<sup>+</sup> ion) of 3892.64 Da, matching native Hm1b.

**Hm1a synthesis.** Solvents for RP-HPLC consisted of 0.05% TFA/H<sub>2</sub>O (solvent A) and 90% MeCN/0.043% trifluoroacetic acid (TFA)/H<sub>2</sub>O (solvent B). Analytical HPLC was performed on a Shimadzu LC20AT system using a Thermo Hypersil GOLD 2.1  $\times$  100 mm C18 column heated at 40°C with a flow rate of 0.3 ml min<sup>-1</sup>. A gradient of 10 to 55% B over 30 min was used, with detection at 214 nm.

Preparative HPLC was performed on a Vydac 218TP1022 column running at a flow rate of 16 ml min<sup>-1</sup> using a gradient of 10 to 50% solvent B over 40 min. Mass spectrometry was performed on an API2000 (ABI Sciex) mass spectrometer in positive ion mode. All reagents were obtained commercially and were used without further purification.

Hm1a was synthesized using regioselective disulfide-bond formation<sup>51–53</sup>. The peptide was assembled on a 0.1-mmol scale using a Symphony (Protein Technologies Inc.) automated peptide synthesizer and a H-Ser(tBu)-2-ClTrt (loading 0.69 mmol g<sup>-1</sup>) polystyrene resin. Couplings were performed in dimethylformamide (DMF) using 5 equivalents of Fmoc-amino acid/(2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU)/N,N-diisopropylethylamine (DIEA) (1:1:1) relative to resin loading for 2  $\times$  20 min. Fmoc deprotection was achieved using 30% piperidine/DMF (1  $\times$  1.5 min, then 1  $\times$  4 min). Non-cysteine amino acid side-chains were protected as Asp(OtBu), Arg(Pbf), Glu(OtBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc) and Tyr(tBu). The cysteine side chains were protected as Cys2,Cys16(Meb), Cys9,Cys21(Dpm), and Cys15,Cys28(Trt). Cleavage from the resin was achieved by treatment with 10% acetic acid/10% trifluoroethanol (TFE)/dichloromethane (DCM) at room temperature for 1 h. The product was precipitated and washed with n-hexane then lyophilized from 1,4-dioxane/MeCN/H<sub>2</sub>O.

The first disulfide bond (Cys15–Cys28) was formed by dissolving the crude product in HFIP (5 ml) and adding it dropwise to a stirred solution of I<sub>2</sub> (4 equivalents) in 10% 1,1,1,3,3,3-hexafluoroisopropanol-2-ol (HFIP)/DCM (20 ml) over 5 min. Stirring was continued for a further 5 min then the solution was poured into a solution of ascorbic acid/sodium acetate in H<sub>2</sub>O. The aqueous phase was extracted with DCM, and the combined organic layers washed with water. Following removal of solvent under reduced pressure, the product was lyophilized from 1,4-dioxane/MeCN/H<sub>2</sub>O. Electrospray ionization mass spectrometry (ESI-MS) (*m/z*): calc. (avg) 2,159.4 [M+3H]<sup>3+</sup>, found 2,159.7.

The remaining side-chain-protecting groups (except Cys(Meb)) were removed by treatment with 95% TFA/2.5% triisopropylsilane (TIPS)/2.5% H<sub>2</sub>O at room temperature for 2 h. After most of the cleavage solution was evaporated under a stream of N<sub>2</sub>, the product was precipitated and washed with cold Et<sub>2</sub>O and lyophilized from 50% MeCN/0.1% TFA/H<sub>2</sub>O to give Cys2,Cys16(Meb), Cys9,Cys21(SH), Cys15–Cys28(SS) Hm1a (280 mg). ESI-MS (*m/z*): calc. (avg) 1,404.3 [M+3H]<sup>3+</sup>, found 1,404.1.

The second disulfide bond (Cys9–Cys21) was formed by dissolving the crude product from the previous step in 30% DMSO/0.1 M HCl (0.5 mg ml<sup>-1</sup>) and stirring at room temperature for 24 h. Cys2,16(Meb), Cys9–Cys21(SS), Cys15–Cys28(SS) Hm1a was then isolated by preparative HPLC (30 mg). ESI-MS (*m/z*): calc. (avg) 1,403.6 [M+3H]<sup>3+</sup>, found 1,403.3.

Formation of the third disulfide bond (Cys2–Cys16) was then achieved by first removing the Cys(Meb) groups by treatment with HF/*p*-cresol (9:1) at 0°C for 1 h. The product was precipitated and washed with cold Et<sub>2</sub>O and lyophilized from 50% MeCN/0.1% TFA/H<sub>2</sub>O yielding Cys2,16(SH), Cys9–Cys21(SS), Cys15–Cys28(SS) Hm1a (24 mg). ESI-MS (*m/z*): calc. (avg) 1,334.1 [M+3H]<sup>3+</sup>, found 1,333.7. Oxidation of the liberated thiols was performed using 30% DMSO/0.1 M HCl as described for the second disulfide bond to yield fully oxidised Hm1a (3 mg) that was indistinguishable by analytical HPLC from an authentic sample. ESI-MS (*m/z*): calc. (avg) 1,333.5 [M+3H]<sup>3+</sup>, found 1,333.1.

**Na<sub>v</sub> and K<sub>v</sub> channel constructs.** Human (h)Na<sub>v</sub>1.4, hNa<sub>v</sub>1.5, and rat (r)K<sub>v</sub>2.1 were provided by P. Ruben, C. Ahern, and K. Swartz, respectively. hNa<sub>v</sub>1.1–1.3, hNa<sub>v</sub>1.6–1.8 were obtained from Origene Technologies, Inc. Accession numbers are NM\_001165963.1 (hNa<sub>v</sub>1.1), NM\_021007.2 (hNa<sub>v</sub>1.2), NM\_006922.3 (hNa<sub>v</sub>1.3), NM\_000334 (hNa<sub>v</sub>1.4), NM\_198056 (hNa<sub>v</sub>1.5), NM\_014191.2 (hNa<sub>v</sub>1.6), NM\_002977.2 (hNa<sub>v</sub>1.7), and NM\_006514.3 (hNa<sub>v</sub>1.8). Channel chimaeras were generated using sequential PCR with rNa<sub>v</sub>1.4 (gift from B. Chanda), K<sub>v</sub>2.1 $\Delta$ 7 (refs 54, 55), hNa<sub>v</sub>1.1, and hNa<sub>v</sub>1.9 (ref. 24) (Origene Technologies: NM\_014139.2) as templates. Mouse K<sub>v</sub>4.1 was obtained from Addgene and originated in the laboratory of L. Salkoff. The K<sub>v</sub>2.1 $\Delta$ 7 construct contains seven point mutations in the outer vestibule that render the channel sensitive to agitoxin-2, a pore-blocking scorpion toxin<sup>56</sup>. cRNA of all constructs was synthesized using T3 or T7 polymerase (mMessage mMachine kit, Life Technologies) after linearizing the fully-sequenced DNA with appropriate restriction enzymes.

**Xenopus oocytes.** Channels and chimaeras were expressed in *Xenopus laevis* oocytes (animals acquired from *Xenopus* one) that were incubated at 17°C in Barth's medium (88 mM NaCl, 1 mM KCl, 0.33 mM Ca(NO<sub>3</sub>)<sub>2</sub>, 0.41 mM CaCl<sub>2</sub>, 0.82 mM MgSO<sub>4</sub>, 2.4 mM NaHCO<sub>3</sub>, 5 mM HEPES, and 0.1 mg ml<sup>-1</sup> gentamycin; pH 7.6 with NaOH) for 1–4 days after cRNA injection, and then were studied using two-electrode voltage-clamp recording techniques (OC-725C; Warner Instruments or GeneClamp 500B; Axon Instruments) with a 150- $\mu$ l recording chamber or a small volume (<20  $\mu$ l) oocyte perfusion chamber (AutoMate Scientific). Data were filtered at 4 kHz and digitized at 20 kHz using pClamp 10

software (Molecular Devices). Microelectrode resistances were 0.5–1 M $\Omega$  when filled with 3 M KCl. For  $K_v2.1$  and  $K_v4.1$  chimaera experiments, the external recording solution contained (in mM): 50 KCl, 50 NaCl, 5 HEPES, 1 MgCl<sub>2</sub> and 0.3 CaCl<sub>2</sub>, pH 7.6 with NaOH. For  $Na_v$  and  $K_v4.1$  experiments, the external recording solution contained (in mM): 100 NaCl, 5 HEPES, 1 MgCl<sub>2</sub> and 1.8 CaCl<sub>2</sub>, pH 7.6 with NaOH. For  $Na_v$  channel experiments, the external recording solution contained (in mM): 100 NaCl, 5 HEPES, 1 MgCl<sub>2</sub> and 1.8 CaCl<sub>2</sub>, pH 7.6 with NaOH. All experiments were performed at room temperature ( $\sim 22^\circ\text{C}$ ) and toxin samples were diluted in recording solution with 0.1% BSA. Leak and background conductance, identified by blocking the channel with agitoxin-2 or TTX, were subtracted for  $K_v$  or  $Na_v$  channel currents, respectively. Voltage–activation relationships were obtained by measuring tail currents for  $K_v$  channels, or by monitoring steady-state currents and calculating conductance for  $Na_v$  channels. Occupancy of closed or resting channels by toxins was examined using negative holding voltages where open probability was low, and the fraction of unbound channels was estimated using depolarizations that are too weak to open toxin-bound channels. After addition of toxin to the recording chamber, equilibration between toxin and channel was monitored using weak depolarizations elicited at 5–10-s intervals. For all channels, voltage–activation relationships were recorded in the absence and presence of toxin. Off-line data analysis was performed using Clampfit 10 (Molecular Devices) and Origin 7.5 (Originlab).

Multiple protocols were used to probe the biophysical characteristics of the  $Na_v$  channels and chimaeras studied. To determine conductance–voltage and steady-state inactivation relationships, oocytes expressing  $Na_v$  channels were held at  $-90\text{ mV}$  and depolarized in 5-mV steps from  $-90\text{ mV}$  to  $5\text{ mV}$  for 50 ms, immediately followed by a step to  $-15\text{ mV}$  to elicit the maximum available current and after 50 ms, returned to the  $-90\text{ mV}$  holding potential. Peak current generated during the incremental portion of the protocol was used to calculate the conductance–voltage relationship while the peak current during the  $-15\text{ mV}$  step as a function of the earlier voltage step was used to determine the steady-state inactivation relationship. The time constant of fast inactivation was determined by fitting single exponential curves to the  $-15\text{ mV}$  step of the aforementioned protocol. Boltzmann curves were fitted in Clampfit 10 (Molecular Devices) and statistics calculated with Excel or the R statistical package (Student's *t*-test).

**Cultured neurons.** Whole-cell patch clamp of cultured mouse trigeminal neurons was performed as described<sup>57</sup>. Buffer solution contained (in mM) 150 NaCl, 2.8 KCl, 1 MgSO<sub>4</sub>, 10 HEPES, pH 7.4 with NaOH and was perfused with or without toxins/drugs using a SmartSquirt Micro-Perfusion system (AutoMate). For colonic DRG, whole-cell recordings were made from fluorescently labelled thoracolumbar (T10–L1) colonic DRG neurons 20–48 h after plating, using fire-polished glass electrodes with a resistance of 2–5 M $\Omega$ . All recordings were performed at room temperature ( $20\text{--}22^\circ\text{C}$ ). Signals were amplified by using an Axopatch 200A amplifier, digitized with a Digidata 1322A and recorded using pCLAMP 9 software (Molecular Devices). For all DRG neurons the holding potential was  $-70\text{ mV}$ . In current clamp mode a series of depolarizing pulses (500 ms, 10 pA step) were applied from holding potential ( $-70\text{ mV}$ ) and the rheobase (amount of current (pA) required for action potential generation) determined. The number of action potentials at  $2 \times$  rheobase was also determined. Depolarizing pulses were tested in normal external bath solution and following the addition of Hm1a (100 nM). Control solutions and Hm1a were applied with a gravity-driven multi-barrel perfusion system positioned within 1 mm of the neuron under investigation. Intracellular solutions contained (in mM): 135 KCl; 2 MgCl<sub>2</sub>; 2 MgATP; 5 EGTA-Na; 10 HEPES-Na; adjusted to pH 7.4. Extracellular solutions contained (in mM): 140 NaCl; 4 KCl; 2 MgCl<sub>2</sub>; 2 CaCl<sub>2</sub>; 10 HEPES-Na; 5 glucose; adjusted to pH 7.4.

**Skin-nerve recordings.** To assess primary afferent activity in response to the Hm1a spider toxin, we used the *ex vivo* skin-nerve preparation, as previously described<sup>58</sup>. Briefly, animals were lightly anaesthetized with inhaled isoflurane and then killed by cervical dislocation. The hair on the lower extremities was shaved, and the hairy skin of the hindpaw was then quickly dissected along with its innervating saphenous nerve. The skin and nerve were then placed in a recording chamber filled with warmed ( $32^\circ\text{C}$ ), oxygenated buffer consisting of (in mM): 123 NaCl, 3.5 KCl, 2.0 CaCl<sub>2</sub>, 1.7 NaH<sub>2</sub>PO<sub>4</sub>, 0.7 MgSO<sub>4</sub>, 9.5 sodium gluconate, 5.5 glucose, 7.5 sucrose and 10 HEPES titrated to a pH of  $7.45 \pm 0.05$ .

The saphenous nerve was then threaded into a mineral oil-filled chamber, teased apart atop an elevated mirror plate, and placed on an extracellular recording electrode. Single-unit receptive fields were then identified with a mechanical search stimulus using a blunt glass probe. A $\delta$  afferents were identified based on a conduction velocity between 1.2 and  $10\text{ m s}^{-1}$ , and were subtyped into A-mechanonociceptors based on their slow adaptation to a mechanical stimulus<sup>59</sup>.

After locating an A-mechanonociceptor fibre, to determine the threshold force for action potential generation, the receptive field was stimulated with calibrated Von Frey filaments. A metal moat (inner diameter: 4.7 mm) was then placed over the centre of the receptive field to isolate it from the surrounding buffer. Buffer

within the moat was then evacuated and replaced with a buffer containing either  $1\text{ }\mu\text{M}$  Hm1a or vehicle (buffer). Receptive fields were incubated with toxin or buffer for 2–5 min. A custom-built, feedback-controlled mechanical stimulator was then placed within the moat and the receptive field was mechanically stimulated with a series of increasing forces (15 mN, 50 mN and 100 mN) for 10 s each. To avoid sensitization/desensitization, a rest period of 1 min was introduced between stimulations.

Data were digitized using a PowerLab A/D converter and recorded using LabChart software and Spike Histogram extension (AD Instruments). All skin-nerve data were recorded and analysed with the experimenter blinded to whether toxin or vehicle was used. Recordings were included in the final data set only if the firing of the fibre was clearly distinguishable from both background noise and any other fibres firing during stimulation.

**Gut-nerve recordings.** *Ex vivo* single-unit extracellular recordings of action potential discharge were made from splanchnic colonic afferents. Recordings were made from healthy or CVH mice using standard protocols<sup>60–62</sup>. Baseline mechanosensitivity was determined in response to application of a 2-g Von Frey hair probe to the afferent receptive field for 3 s. This process was repeated 3–4 times, separated each time by 10 s. Mechanosensitivity was then re-tested after the application of Hm1a (100 nM) or the  $Na_v1.1$  blocker ICA-121432 (500 nM) or a combination of both ICA-121432 (500 nM) and Hm1a (100 nM). Instantaneous frequency is defined as the inverse of the time interval between an action potential and the previous action potential. After application of Hm1a, significant increases ( $P < 0.05$ ) in mechanically evoked firing were seen overall in both healthy and CVH fibres. However, in both conditions we clearly recorded Hm1a-responsive and Hm1a-non-responsive neurons. We therefore binned fibres by responsiveness and present these data separately for clarity (Fig. 5 and Extended Data Fig. 5). Group data are presented as spikes per second and are expressed as mean  $\pm$  s.e.m.

**Animal use, husbandry and genotyping.** Mice were bred and housed in accordance with UCSF Institutional Animal Care Committee (IACUC) guidelines. Animals were housed in groups of 2–5 with constant access to food and water. Floxed *Scn1a* mice<sup>13</sup> were provided by W. Catterall. Floxed mice were bred to peripherin-Cre (Per-Cre) mice<sup>36</sup> to produce *Scn1a*<sup>fl/fl</sup>  $\times$  Per-Cre conditional knockout mice.  $Na_v1.1$  floxed alleles were detected using primers previously described<sup>13</sup> and Per-Cre expression was detected using the following primers to Cre recombinase: Cre\_F: TAGCGTTCGAACGCACTGATTTCG; Cre\_R: CGCCGTAATCAATCGATGAGTTG.

Somatic behavioural experiments were approved by UCSF IACUC and were in accordance with the National Institutes of Health (NIH) Guide of the Care and Use of Laboratory Animals and the recommendation of the International Association for the Study of Pain. Animals used in skin-nerve recordings were naive C57BL/6 male mice ( $n = 10$ ), aged 6–16 weeks. Mice were housed on a 14:10 h light:dark cycle with ad libitum access to food and water in a climate-controlled room. All protocols were approved by the Institutional Animal Care and Use Committee at the Medical College of Wisconsin. Animals used in colonic afferent and colonic DRG neuron studies were male C57BL/6J mice. The Animal Ethics Committees of The University of Adelaide and the South Australian Health and Medical Research Institute (SAHMRI) approved experiments involving animals.

**Sensory neuron culture and calcium imaging.** Trigeminal ganglia were dissected from newborn (P0–P3) Sprague–Dawley rats or C57BL/6 mice and cultured for  $>12\text{ h}$  before calcium imaging or electrophysiological recording. Embryonic DRG cultures were provided by J. Chan<sup>63</sup>. Embryonic cultures were maintained as described and calcium imaging experiments were performed 1–10 d after primary cultures were established. Primary cells were plated onto cover slips coated with poly-L-lysine (Sigma) and laminin (Invitrogen,  $10\text{ }\mu\text{g ml}^{-1}$ ). Cells were loaded for calcium imaging with Fura-2-AM (Molecular Probes) for  $>1\text{ h}$ . Buffer solution ((in mM), 150 NaCl, 2.8 KCl, 1 MgSO<sub>4</sub>, 10 HEPES, pH 7.4 with NaOH) was perfused with or without toxins/drugs using a SmartSquirt Micro-Perfusion system (AutoMate).

**In situ hybridization and immunohistochemistry.** *In situ* hybridization (ISH) was performed using ViewRNA ISH Tissue 2-plex or 1-plex Assay Kits (Affymetrix). Target mRNA signals appear as puncta in bright field or fluorescent microscopy. Eight-to-twelve-week-old mice were deeply anaesthetized with pentobarbital then transcardially perfused with 10 ml PBS followed by 10 ml 10% neutral buffered formalin (NBF). DRGs were dissected, post-fixed in 10% NBF at  $4^\circ\text{C}$  overnight, cryoprotected in PBS with 30% (w/v) sucrose overnight at  $4^\circ\text{C}$ , then embedded in OCT compound at  $-20^\circ\text{C}$ . Tissue was sectioned at  $12\text{ }\mu\text{m}$ , thaw-captured on Diamond White Glass slides (Globe Scientific), and stored at  $-20^\circ\text{C}$  until use. Slides were used within 2 weeks of processing to produce optimal signals.

ViewRNA ISH Tissue 2-plex assay was performed with frozen tissue modifications as indicated by the manufacturer including the endogenous alkaline phosphatase inactivation by incubation in H<sub>2</sub>O with 0.1 M HCl and 300 mM NaCl. The haematoxylin and eosin counterstaining procedure was omitted. Images were

acquired with a Leica DMRB microscope and DFC500 digital camera using Leica Application Suite v3.5.0 then analysed further using ImageJ software.

To co-label neuronal subpopulations markers (NF200, IB4, CGRP, TH) and *Na<sub>v</sub>1.1* mRNA, ViewRNA ISH Tissue 1-plex assay and immunohistochemistry (IHC) were performed sequentially using a protocol modified from ref. 64. ISH/IHC was not found to be compatible with all primary antibodies. Animals, tissue, and slides were prepared as described in the preceding paragraph. Frozen slides with tissue sections were warmed in a vacuum oven for 10 min at 60°C, fixed in PBS with 4% (v/v) formaldehyde for 10 min at room temperature and then processed according to the manufacturer's protocol with frozen tissue modifications in a ThermoBrite Slide Processing System (Abbott Molecular). Washing steps were performed as indicated, in a deliberate and vigorous manner. Optimal protease and probe incubation times were determined to be 12 min and 2 h, respectively. Following development in Fast Red Substrate, slides were rinsed briefly in PBS then immediately processed for IHC. Slides were incubated for 1 h at room temperature in a blocking solution consisting of PBS with 0.1% (v/v) Triton X-100 (Sigma) and 10% normal goat serum (NGS). Slides were then incubated in primary antibody solution (PBS with 0.1% Triton X-100 and 2.5% NGS) overnight at 4°C, vigorously agitated for 2 min in fresh PBS three times, then incubated in secondary antibody solution (PBS with 0.1% (v/v) Triton X-100) for 2 h at room temperature in the dark. Sections were then washed by vigorous agitation for 2 min in fresh PBS three times before mounting with ProLong Gold antifade reagent with DAPI (Life Technologies) and addition of coverslips. Images were acquired with a Leica DMRB microscope and DFC500 digital camera using Leica Application Suite v3.5.0 then further analysed using ImageJ software.

Affymetrix was commissioned to design a Type 1 probe set to mouse *Na<sub>v</sub>1.1* (*Scn1a*, nM\_018733.2) and Type 6 probe sets to mouse TRPV1 (*Trpv1*, nM\_001001445.2), mouse *Na<sub>v</sub>1.7* (*Scn9a*, nM\_001290674.1), mouse 5HT<sub>3</sub> (*Htr3a*, nM\_001099644.1), and mouse TRPM8 (*Trpm8*, nM\_134252.3) coding regions. We used the following primary antibodies: mouse anti-NF200 (1:10,000, Sigma), rabbit anti-CGRP (1:10,000, Peninsula Labs), and rabbit anti-TH (1:5,000, AbCam). We used fluorophore-conjugated secondary antibodies raised in goat against mouse or rabbit, as appropriate (1:1,000, Alexa Fluor 488, Life Technologies). To identify IB4-binding cells, biotinylated IB4 (1:1,000, Vector Labs) and fluorophore-conjugated streptavidin (1:1,000, Alexa Fluor 488, Life Technologies) were used in place of primary and secondary antibodies. Fos staining was performed 90 min after hindpaw injection of Hm1a or PBS. Spinal cord sections were prepared from lumbar L4/L5 and stained with rabbit anti-Fos (1:5,000, CalBiochem). ATF3 antibody (Santa Cruz Biotechnology) was used at 1:2,000.

**Statistics and experimental design.** Sample sizes for cellular physiology, histology and animal behaviour were chosen based on previous experience with these assays as the minimum number of independent observations required for statistically significant results. No statistical methods were used to predetermine sample size. For histology, at least three sections from each of at least three animals were counted. For oocyte and mouse neuron experiments, multiple batches or litters were used for all experiments. For behavioural experiments, animals were randomly chosen for different experimental cohorts by a blinded experimenter. Experimental and control conditions were compared within the same experimental time-course using randomly selected animals from one or multiple cages. Responses were then scored by an experimenter blinded to injection condition and experimental cohort. Animal genotype was tracked by ear tags and genotype unblinding occurred after analysis was complete.

Data were analysed using Prism 6 software (GraphPad Software) and significance testing used either Student's *t*-tests or one-way ANOVA followed by Bonferroni or Tukey's post hoc tests, as noted in figure legends. All significance tests are two-sided. The number of experiments (*n*) and significance are reported in the figure legends. All significance tests were justified as appropriate given the experimental design and nature of the comparisons. We assume equal variance and normally distributed data within experimental paradigms where comparisons are made. These are common assumptions relied upon for significance testing within these experimental paradigms as previously published by our group and others.

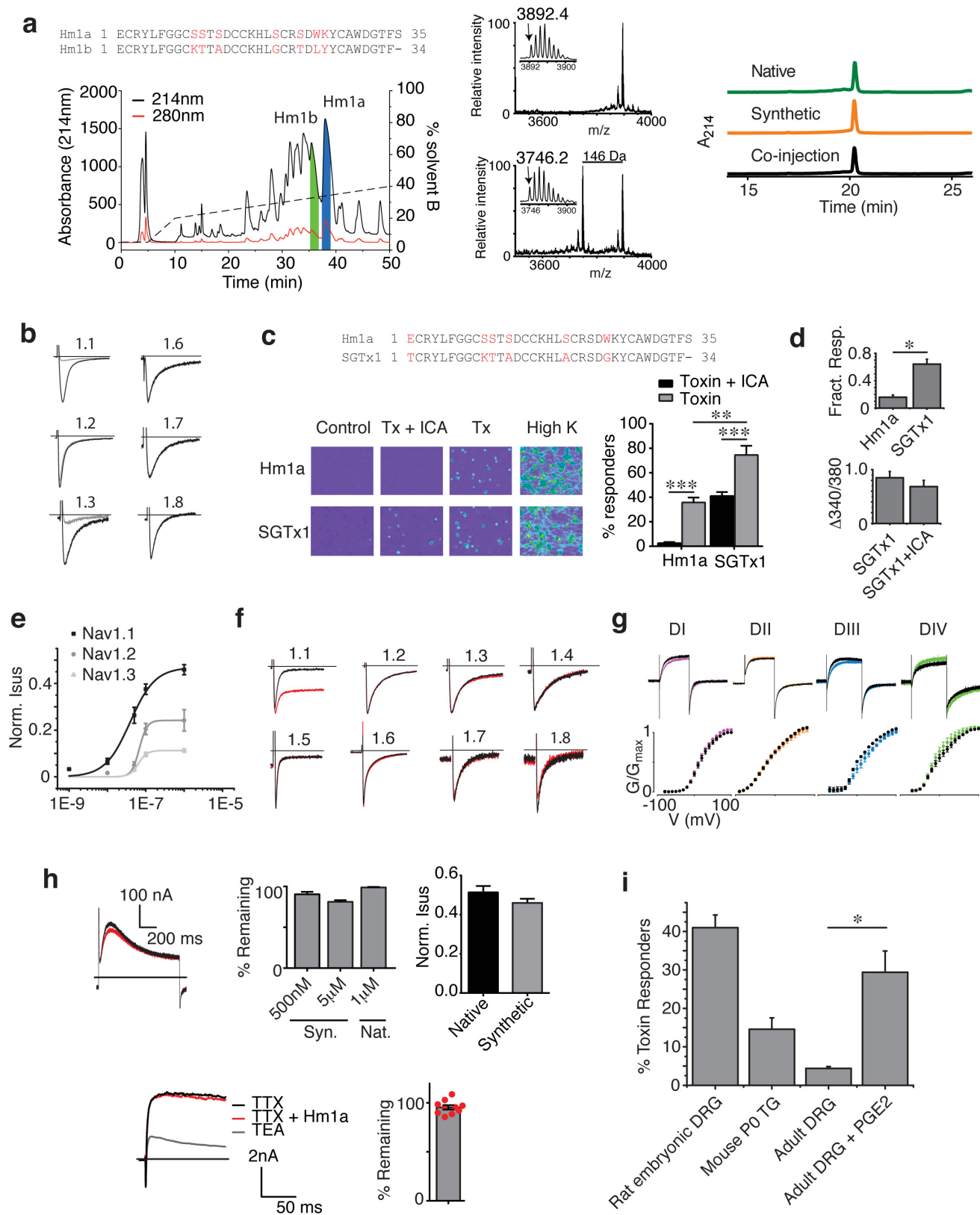
**Behaviour.** For behavioural experiments in Fig. 4, adult mice (6–12 weeks) were used and randomly selected for analysis. Male and female mice were first considered separately in hindpaw nocifensive response experiments. Both sexes showed significantly greater responses to toxin in wild-type littermate versus *Na<sub>v</sub>1.1*<sup>F/F</sup> × Per-Cre CKO mice (one-sided unpaired Student's *t*-test, *P* < 0.05, wild-type female: *n* = 5, CKO female: *n* = 6, wild-type male: *n* = 5, CKO male: *n* = 5). Therefore, male and female behavioural responses were pooled and subsequent experiments were performed on both male and female mice for CKO and wild-type littermate experiments, or only male mice for other conditions (for example, capsaicin ablation). Nocifensive responses were recorded during a 20-min observation period immediately following intraplantar injections (10 μl PBS with or without 5 μM Hm1a). Licking/biting behaviour was scored as seconds of behaviour with the experimenter

blinded to injection condition and experimental cohort (wild-type, CKO or capsaicin-ablated mice). Hargreaves and Von Frey tests were performed 30 min after intraplantar injection of 500 nM Hm1a or Hm1b to measure heat and mechanical sensitivity, respectively. Intrathecal (i.t.) capsaicin ablation was performed as previously described<sup>33</sup>, and i.t. capsaicin-treated mice were tested on a hot plate to ensure ablation of TRPV1<sup>+</sup> afferents. Ablation was also confirmed by histology.

**Model of chronic visceral hypersensitivity.** Colitis was induced by administration of 2,4,6-trinitrobenzenesulfonic acid (TNBS) as described previously<sup>61,62</sup>. Briefly, 13-week-old anaesthetized mice were administered an intra-colonic enema of 0.1 ml TNBS (130 μg ml<sup>-1</sup> in 30% ethanol) via a polyethylene catheter<sup>61,62,65</sup>. Histological examination of mucosal architecture, cellular infiltrate, crypt abscesses, and goblet cell depletion confirmed significant TNBS-induced damage by day 3 post-treatment, which largely recovered by day 7, and fully recovered by 28 days. High-threshold nociceptors from mice at the 28-day time point displayed significant mechanical hypersensitivity, lower mechanical activation thresholds, and hyperalgesia and allodynia<sup>40</sup>. As such, they are termed chronic visceral hypersensitivity (CVH) mice<sup>61,62,65,66</sup>.

**Retrograde tracing and cell culture of colonic DRG neurons.** Healthy and CVH mice of 16 weeks of age were anaesthetized with halothane and following midline laparotomy, three 10-μl injections of the fluorescent retrograde neuronal tracer cholera toxin subunit B conjugated to Alexa Fluor 488 were made sub-serosally within the wall of the descending colon. Four days after injection, mice were killed by CO<sub>2</sub> inhalation and DRGs from T10–L1 were surgically removed. DRGs were digested with 4 mg ml<sup>-1</sup> collagenase II (GIBCO, Invitrogen) and 4 mg ml<sup>-1</sup> dispase (GIBCO) for 30 min at 37°C, followed by 4 mg ml<sup>-1</sup> collagenase II for 10 min at 37°C. Neurons were mechanically dissociated into a single-cell suspension via trituration through fire-polished Pasteur pipettes. Neurons were resuspended in DMEM (GIBCO) containing 10% FCS (Invitrogen), 2 mM L-glutamine (GIBCO), 100 μM MEM non-essential amino acids (GIBCO) and 100 mg ml<sup>-1</sup> penicillin/streptomycin (Invitrogen). Neurons were spot-plated on 8-mm HCl-treated coverslips coated with poly-D-lysine (800 μg ml<sup>-1</sup>) and laminin (20 μg ml<sup>-1</sup>) and maintained in an incubator at 37°C in 5% CO<sub>2</sub>.

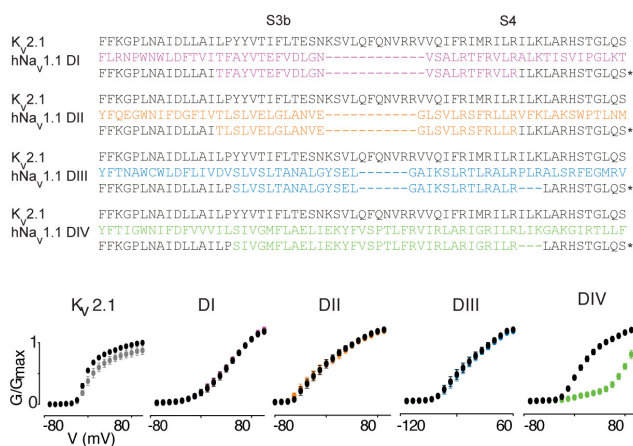
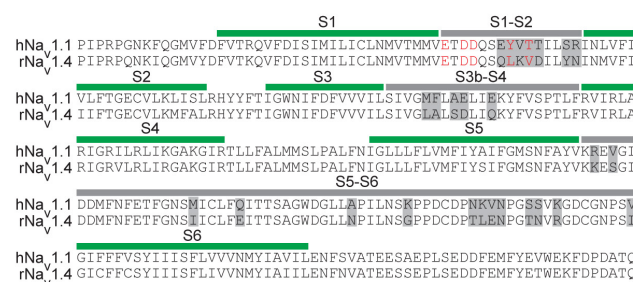
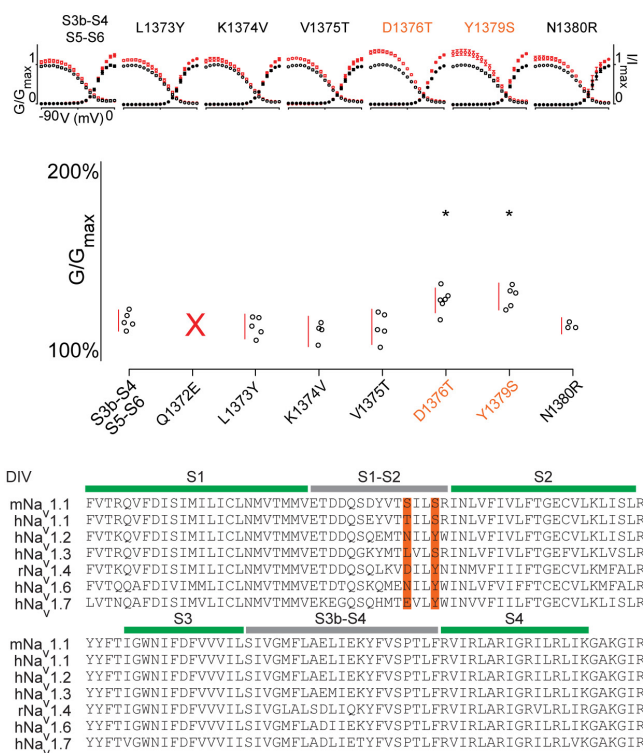
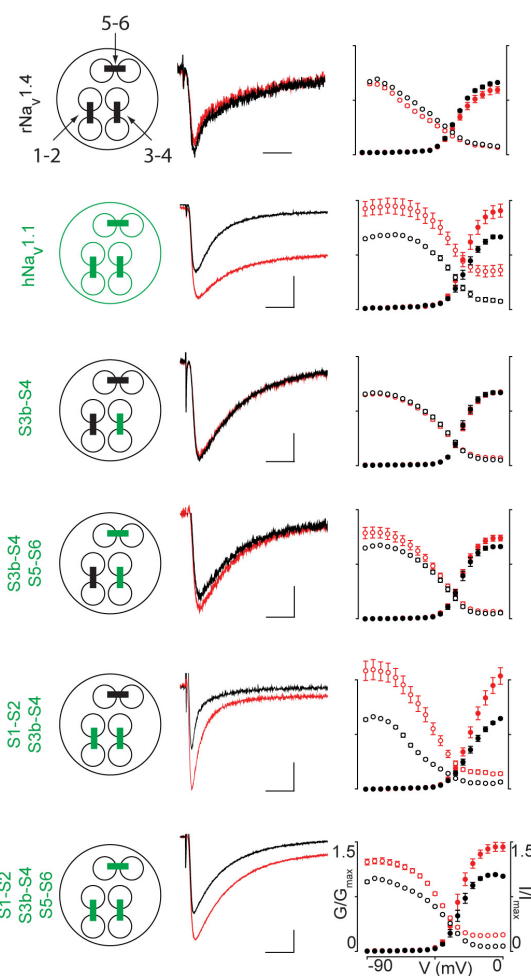
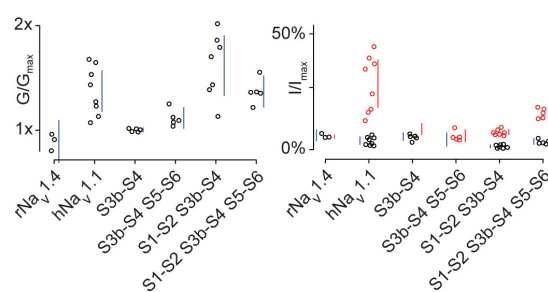
50. Chow, C. Y., Cristofori-Armstrong, B., Undheim, E. A. B., King, G. F. & Rash, L. D. Three peptide modulators of the human voltage-gated sodium channel 1.7, an important analgesic target, from the venom of an Australian tarantula. *Toxins (Basel)* **7**, 2494–2513 (2015).
51. Kamber, B. *et al.* The synthesis of cystine peptides by iodine oxidation of S-trityl-cysteine and S-acetamidomethyl-cysteine peptides. *Helv. Chim. Acta* **63**, 899–915 (1980).
52. Vetter, I. *et al.* Isolation, characterization and total regioselective synthesis of the novel μO-conotoxin MfVIA from *Conus magnificus* that targets voltage-gated sodium channels. *Biochem. Pharmacol.* **84**, 540–548 (2012).
53. Góngora-Benítez, M. *et al.* Acid-labile Cys-protecting groups for the Fmoc/tBu strategy: filling the gap. *Org. Lett.* **14**, 5472–5475 (2012).
54. Frech, G. C., Van Dongen, A. M., Schuster, G., Brown, A. M. & Joho, R. H. A novel potassium channel with delayed rectifier properties isolated from rat brain by expression cloning. *Nature* **340**, 642–645 (1989).
55. Swartz, K. J. & MacKinnon, R. Hanatoxin modifies the gating of a voltage-dependent K<sup>+</sup> channel through multiple binding sites. *Neuron* **18**, 665–673 (1997).
56. Garcia, M. L., Garcia-Calvo, M., Hidalgo, P., Lee, A. & MacKinnon, R. Purification and characterization of three inhibitors of voltage-dependent K<sup>+</sup> channels from *Leiurus quinquestriatus* var. *hebraeus* venom. *Biochemistry* **33**, 6834–6839 (1994).
57. Bohlen, C. J. *et al.* A bivalent tarantula toxin activates the capsaicin receptor, TRPV1, by targeting the outer pore domain. *Cell* **141**, 834–845 (2010).
58. Reeh, P. W. Sensory receptors in mammalian skin in an in vitro preparation. *Neurosci. Lett.* **66**, 141–146 (1986).
59. Koltzenburg, M., Stucky, C. L. & Lewin, G. R. Receptive properties of mouse sensory neurons innervating hairy skin. *J. Neurophysiol.* **78**, 1841–1850 (1997).
60. Brierley, S. M., Jones, R. C., III, Gebhart, G. F. & Blackshaw, L. A. Splanchnic and pelvic mechanosensory afferents signal different qualities of colonic stimuli in mice. *Gastroenterology* **127**, 166–178 (2004).
61. Castro, J. *et al.* Linacotide inhibits colonic nociceptors and relieves abdominal pain via guanylate cyclase-C and extracellular cyclic guanosine 3',5'-monophosphate. *Gastroenterology* **145**, 1334–1346 (2013).
62. de Araujo, A. D. *et al.* Selenoether oxytocin analogues have analgesic properties in a mouse model of chronic abdominal pain. *Nat. Commun.* **5**, 3165 (2014).
63. Lewallen, K. A. *et al.* Assessing the role of the cadherin/catenin complex at the Schwann cell-axon interface and in the initiation of myelination. *J. Neurosci.* **31**, 3032–3043 (2011).
64. Bardoni, R. *et al.* Delta opioid receptors presynaptically regulate cutaneous mechanosensory neuron input to the spinal cord dorsal horn. *Neuron* **81**, 1312–1327 (2014).
65. Hughes, P. A. *et al.* Post-inflammatory colonic afferent sensitization: different subtypes, different pathways and different time courses. *Gut* **58**, 1333–1341 (2009).
66. Hughes, P. A. *et al.* Increased κ-opioid receptor expression and function during chronic visceral hypersensitivity. *Gut* **63**, 1199–1200 (2014).



Extended Data Figure 1 | See next page for caption.

**Extended Data Figure 1 | Hm1a and Hm1b selectively target  $\text{Na}_v1.1$  in sensory neurons.** **a**, Left, HPLC chromatogram showing reversed-phase C18 fractionation of *Heteroscodra maculata* venom; peaks containing Hm1a and Hm1b are labelled. Peptide sequences as determined by Edman degradation are displayed above. Middle, MALDI-TOF spectra of native undigested Hm1b (top) and native Hm1b digested with carboxypeptidase Y for 20 min (bottom), with inserted spectra showing the monoisotopic mass of each in daltons (Da). The observed mass difference of 146 Da between the intact and digested Hm1b corresponds to the final residue, Phe, with an amidated C terminus. Right, chromatograms show reversed-phase C18 HPLC profiles of native and synthetic Hm1a, which were indistinguishable when co-injected. **b**, Representative currents from oocytes expressing hNa<sub>v</sub> subtypes before (black) and after (grey) bath application of ICA-121431 (500 nM). Currents were elicited during 1-Hz stimulation to induce use-dependent block. **c**, Top, amino acid sequence comparison of Hm1a with SGTx1, a related, but non-selective fast-inactivation inhibitor. Bottom, representative calcium imaging experiment comparing ICA-121431-mediated block of Hm1a- or SGTx1-evoked responses in cultured embryonic DRG neurons, with group data at right (\*\* $P < 0.01$ , \*\*\* $P < 0.001$ ,  $n = 4$ ). **d**, Top, fraction of P0 mouse neurons responding to Hm1a versus SGTx1 (\*\* $P < 0.01$ ). Bottom, ratiometric calcium responses elicited by SGTx1 (500 nM) in the presence and absence

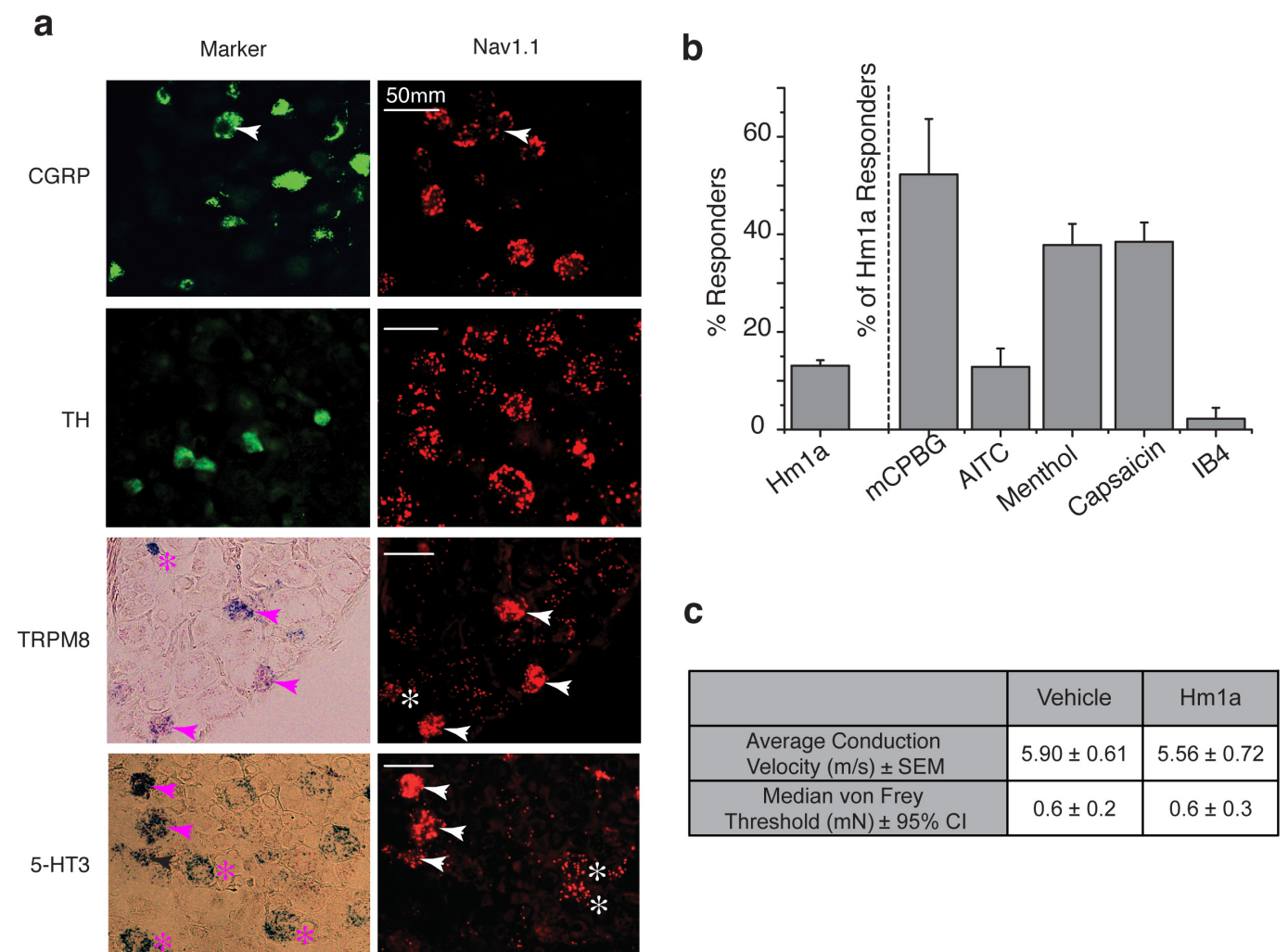
of ICA-121431 (500 nM). **e**, Dose-response curves for Hm1a inhibition of fast inactivation in oocytes expressing Na<sub>v</sub>1.1, Na<sub>v</sub>1.2 or Na<sub>v</sub>1.3. Sustained current at the end of a 100-ms pulse is normalized to peak current to quantify magnitude of the effect. EC<sub>50</sub> values for hNa<sub>v</sub>1.1 = 38 nM, hNa<sub>v</sub>1.2 = 236 nM and hNa<sub>v</sub>1.3 = 220 nM. **f**, Representative traces from oocytes expressing hNa<sub>v</sub> subtypes in response to a saturating dose (on hNa<sub>v</sub>1.1) of purified Hm1b during a 100-ms depolarization. **g**, rK<sub>v</sub>2.1 chimaeras containing different Na<sub>v</sub>1.9 S3b-S4 motifs were tested for sensitivity to hHm1a (100 nM). Representative traces (top) and summary data (bottom) show a lack of toxin sensitivity for each chimaera. **h**, Top left, representative currents from oocytes expressing mK<sub>v</sub>4.1 before (black) and after (red) bath application of Hm1a (5  $\mu\text{M}$ ). Middle, quantification of mK<sub>v</sub>4.1 blockade by synthetic or native Hm1a. Top right, comparison of sustained current during application of native or synthetic Hm1a (1  $\mu\text{M}$ ) shows similar effects on Na<sub>v</sub>1.1 inactivation. Bottom, representative traces (left) showing that outward currents in P0 trigeminal mouse neurons are unaffected by Hm1a (500 nM). Scatter plot (right,  $n = 10$ ) shows no significant difference. **i**, Percentage of Hm1a (500 nM)-responsive neurons in various culture conditions as assessed by calcium imaging ( $n = 3-4$ , \* $P < 0.05$ ). Error bars represent mean  $\pm$  s.e.m.  $P$  values based on two-way ANOVA with post hoc Tukey's test (c) or unpaired two-tailed Student's  $t$ -test (d, i).

**a****b****c****d****e**

Extended Data Figure 2 | See next page for caption.

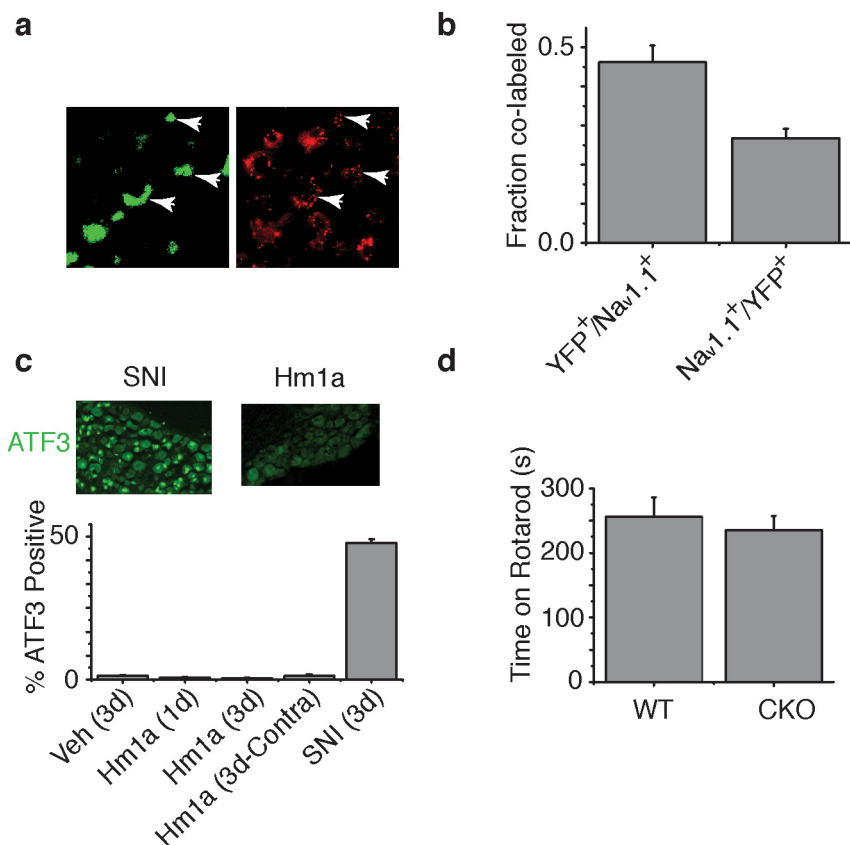
**Extended Data Figure 2 | Hm1a selectivity depends on DIV S1–S2 and S3b–S4 regions.** **a**, Top, alignments between K<sub>v</sub>2.1 and hNa<sub>v</sub>1.1 S3b–S4 regions from each domain (as indicated) with sequence of chimaeras shown below each alignment. Bottom, *G*–*V* relationships from chimaeric channels expressed in oocytes in the absence (black) and presence (colours) of Hm1a (100 nM). **b**, Sequence alignment of hNa<sub>v</sub>1.1 and rNa<sub>v</sub>1.4 showing putative transmembrane segments (green) and regions swapped in chimaeric channels (grey). **c**, Top, using the background of Na<sub>v</sub>1.4 chimaera containing the S3b–S4 and S5–S6 regions of Na<sub>v</sub>1.1, individual residues were mutated in the S1–S2 loop to the cognate residue in Na<sub>v</sub>1.1. The D1376T and Y1379S point mutants in the chimaeric rNa<sub>v</sub>1.4 channel reveal an increase in peak current after 100 nM Hm1a application (red) relative to untreated controls (black). Filled circles denote *G*–*V* relationships, where oocytes were depolarized for 50 ms in 5-mV steps from a holding potential of –90 mV. Open circles denote steady-state inactivation (*I*/*I*<sub>max</sub>) relationships, where oocytes were depolarized from –90 mV to +5 mV in 5-mV increments for 50 ms preceding a 50-ms step to –15 mV. Middle, dot plot detailing per cent increase in peak conductance of each point mutant in response to 100 nM Hm1a treatment. Each point represents a single oocyte; red bars indicate 95% confidence interval. Mutations highlighted in orange (D1376T and Y1379S) are

statistically different from S3b–S4/S5–S6 control (\**P* < 0.01, Student's *t*-test). The Q1372E mutant did not generate currents. Bottom, alignment of DIV S1–S4 regions from relevant mouse and human Na<sub>v</sub> isoforms. Orange highlights location of residues in the S1–S2 loop that putatively contribute to the toxin effect. **d**, Left, stylized DIV with transmembrane segments represented as circles and extracellular loops as bars (black for native rNa<sub>v</sub>1.4 channel and green for portions transplanted from hNa<sub>v</sub>1.1). Middle, traces displaying effect of Hm1a on each chimaera depolarized to –15 mV from a holding potential of –90 mV. Right, conductance–voltage (*G*/*G*<sub>max</sub>) and steady-state inactivation (*I*/*I*<sub>max</sub>) relationships of each channel and chimaera before and after toxin (black and red, respectively) across a voltage range spanning –90 mV to 0 mV from a holding potential of –90 mV in 5-mV increments. Scale bars as in Fig. 2. **e**, Dot plots displaying the effect of 100 nM Hm1a on peak current (left) and persistent current (right). Data in the left plot were generated by dividing peak conductance after Hm1a application by the peak conductance before. Right plot shows persistent current divided by peak current before (black) or after (red) toxin addition. Persistent current was determined by averaging current from the final millisecond of depolarization to 0 mV from a holding potential of –90 mV. Vertical bars indicate 95% confidence interval.



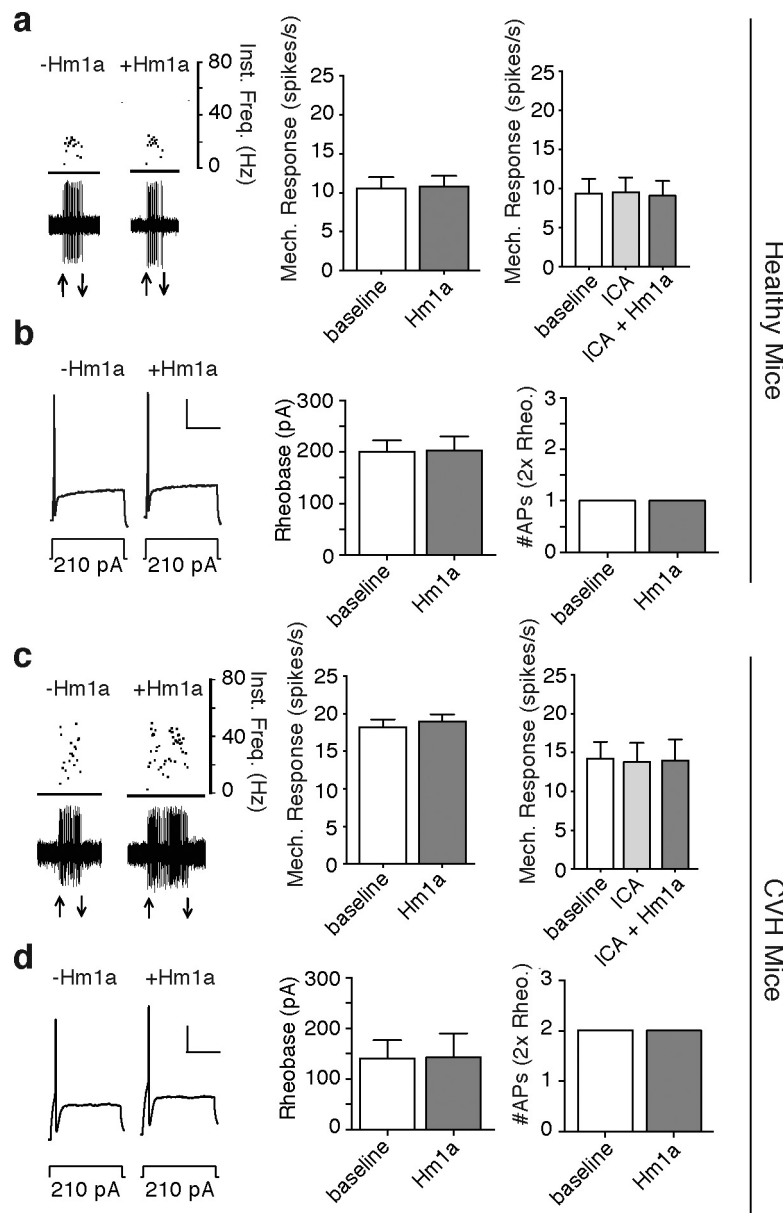
**Extended Data Figure 3 |  $\text{Na}_v1.1$  is expressed by myelinated, non-C-fibre sensory neurons.** **a**, Representative images showing expression of various molecular markers and their overlap with  $\text{Na}_v1.1$  transcripts. Markers include immunohistochemical staining for CGRP and tyrosine hydroxylase (TH) and *in situ* histochemistry for TRPM8 and 5-HT<sub>3</sub> ion channel transcripts. Quantification of overlap for these markers is shown in Fig. 3. **b**, Quantification of the number of toxin-responsive cells in

P0 mouse trigeminal cultures as assessed by calcium imaging (leftmost column) and the percentage of toxin-sensitive cells that responded to other agonists (1-(m-chlorophenyl)-biguanide (mCPBG), allyl isothiocyanate (AITC), capsaicin, and menthol activate 5-HT<sub>3</sub>, TRPA1, TRPV1 and TRPM8 channels, respectively), or bound the lectin IB4. **c**, Table including conduction velocity and Von Frey thresholds for skin-nerve experiments presented in Fig. 3d. Error bars represent mean  $\pm$  s.e.m.



**Extended Data Figure 4 | Control experiments.** These data show control experiments related to Fig. 4. **a**, Representative DRG sections from peripherin-Cre adult mouse showing neurons that express Cre recombinase as visualized using a floxed-stop YFP reporter mouse. *In situ* hybridization histochemistry shows overlap with Nav1.1 transcripts (right). **b**, Quantification of overlap between YFP and Nav1.1. **c**, Comparison of ATF3 induction in DRG following sciatic nerve ligation

(SNI) or intraplantar Hm1a injection. SNI induced ATF3 expression in >50% of DRG neurons whereas ATF3 induction after Hm1a injection was indistinguishable from vehicle (measured 1 or 3 days post-injection). **d**, Peripherin-Cre  $\times$  floxed Nav1.1 mice were compared with wild-type littermates in the rotarod test. No significant differences were observed (unpaired Student's *t*-test). Error bars represent mean  $\pm$  s.e.m.



**Extended Data Figure 5 | A subset of colonic afferents does not express functional  $\text{Na}_v1.1$ .** **a**, Left, representative *ex vivo* colonic single fibre recording from an Hm1a (100 nM)-non-responsive high-threshold fibre from a healthy mouse (arrows indicate application and removal of 2 g Von Frey hair stimulus). Middle, group data showing lack of Hm1a-mediated responses from a subset (9 out of 15) of fibres. Right, group data showing a population (5 out of 10) of healthy, high-threshold mechanoreceptor colonic afferents unaltered by ICA-121432 in the presence or absence of Hm1a (100 nM). **b**, Left, representative whole-cell current clamp recording of a retrogradely traced colonic DRG neuron in response to 500-ms current injection at rheobase. Recordings were made from the same neuron of a healthy control mouse before and after incubation with Hm1a (10 nM). Horizontal scale bar, 250 ms; vertical scale bar, 20 mV.

Middle and right, group data show no effect of Hm1a application on electrical excitability in a sub-population (6 out of 11) of colonic DRG neurons. **c**, Left, representative high-threshold mechanoreceptive colonic fibres from CVH mice showing no change after application of Hm1a (100 nM). Middle, group data from Hm1a-non-responsive colonic fibres (4 out of 11). Right, group data showing a subpopulation of CVH colonic afferents (3 out of 10) unaltered by ICA-121432 in the presence or absence of Hm1a. **d**, Left, representative Hm1a-non-responsive colonic DRG neuron in whole-cell current clamp. Middle and right, group data show electrical excitability is unaltered by Hm1a in a subset (4 out of 11) of CVH colonic DRG neurons. Error bars represent mean  $\pm$  s.e.m. No significant differences were observed using Student's *t*-test (**a–d**, middle; **b**, **d**, right) or one-way ANOVA with post hoc Bonferroni test (**a**, **c**, right).

# Defining the consequences of genetic variation on a proteome-wide scale

Joel M. Chick<sup>1\*</sup>, Steven C. Munger<sup>2\*</sup>, Petr Simecek<sup>2</sup>, Edward L. Huttlin<sup>1</sup>, Kwangbom Choi<sup>2</sup>, Daniel M. Gatti<sup>2</sup>, Narayanan Raghupathy<sup>2</sup>, Karen L. Svenson<sup>2</sup>, Gary A. Churchill<sup>2§</sup> & Steven P. Gygi<sup>1§</sup>

**Genetic variation modulates protein expression through both transcriptional and post-transcriptional mechanisms. To characterize the consequences of natural genetic diversity on the proteome, here we combine a multiplexed, mass spectrometry-based method for protein quantification with an emerging outbred mouse model containing extensive genetic variation from eight inbred founder strains. By measuring genome-wide transcript and protein expression in livers from 192 Diversity Outbred mice, we identify 2,866 protein quantitative trait loci (pQTL) with twice as many local as distant genetic variants. These data support distinct transcriptional and post-transcriptional models underlying the observed pQTL effects. Using a sensitive approach to mediation analysis, we often identified a second protein or transcript as the causal mediator of distant pQTL. Our analysis reveals an extensive network of direct protein-protein interactions. Finally, we show that local genotype can provide accurate predictions of protein abundance in an independent cohort of Collaborative Cross mice.**

Regulation of protein abundance is vital to cellular functions and environmental response. According to the central dogma<sup>1</sup>, the coding sequence of DNA is transcribed into mRNA (transcript), which in turn is translated into protein. Although rates of transcription, translation and degradation of both transcript and protein vary, under this simplest model of regulation, the cellular pool of a protein is determined by the abundance of its corresponding transcript. Genetic or environmental perturbations that alter transcription would directly affect protein abundance. In reality, many layers of regulation intervene in this process, and numerous studies have been carried out to determine whether and to what extent transcript abundance is a predictor of protein abundance<sup>2–6</sup>. Several studies have reported that there is generally a low correlation between the two. An emerging consensus is that much of the protein constituent of the cell is buffered against transcriptional variation<sup>4,7</sup>, but a global perspective of protein buffering has not been put forward.

Genetic variants can influence transcript and protein levels in a quantitative manner. Mapping quantitative trait loci (QTL) that affect transcript (eQTL) or protein (pQTL) abundance in model organisms or human cell lines can identify causal variants and provide a tool to dissect the mode of regulation of gene expression<sup>8</sup>. Analyses of eQTL have yielded a global but incomplete understanding of the regulatory mechanisms associated with gene expression<sup>9–13</sup>. Until now, pQTL analysis has been applied to a modest set of proteins through shotgun proteomics or targeted protein analysis<sup>5,7,14–19</sup>. Much of the pioneering work behind pQTL analysis has been conducted in yeast crosses using mass spectrometry<sup>14–16</sup> or green fluorescent protein (GFP) fusions<sup>20</sup>. Recent advances in quantitative proteomics<sup>21,22</sup> present the possibility of near-comprehensive, genome-wide pQTL analysis.

To investigate how genetic variation affects transcript and protein abundance globally requires a broad set of perturbations. The Diversity Outbred (DO) mouse model is a heterogeneous stock derived from the same eight founder strains as the Collaborative Cross (CC) mice<sup>23–25</sup> (Fig. 1a). The founder strains are fully sequenced<sup>26</sup> and capture a considerable cross-section of the genetic variation present in laboratory

and wild mouse populations. The balanced allele frequencies and simple population structure of the DO mice provides high power and precision for mapping QTL with relatively small sample sizes relative to human mapping studies. We designed a QTL mapping approach that takes advantage of these unique properties of the DO and our knowledge of the founder genomes<sup>27</sup>. For each individual DO mouse, we imputed the founder strain ancestry at 64,000 evenly spaced loci across the diploid genome.

## Gene and protein expression profiling

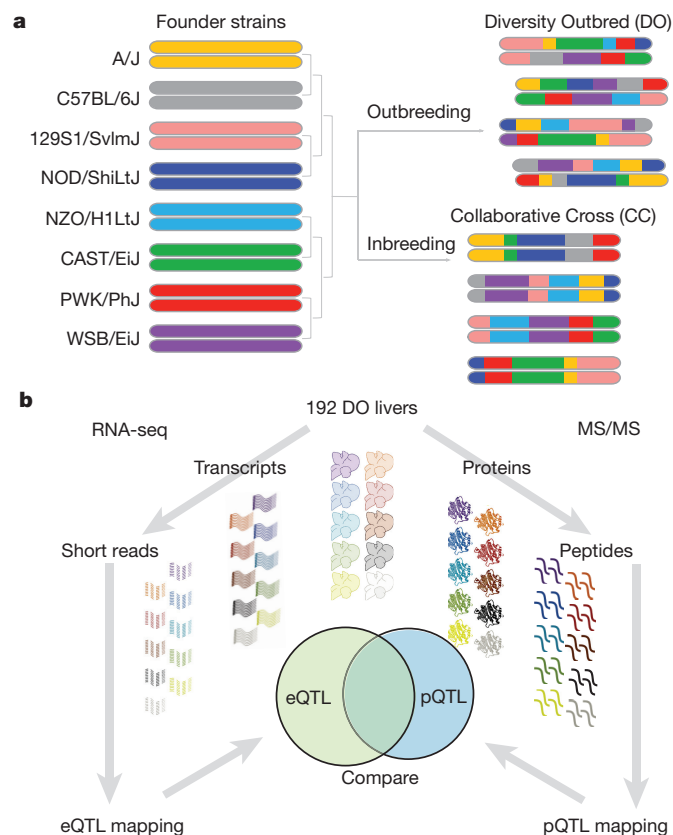
We first applied multiplexed proteomics to evaluate the extent of protein abundance variation among the eight DO/CC founder strains. Founder strain liver proteins were analysed in duplicate from both sexes (Extended Data Fig. 1a, Supplementary Table 1). Protein abundance was highly variable across the eight founder strains; hierarchical clustering and principal component analysis suggested that strain was the major factor driving variation followed by sex. This analysis confirmed our expectation that the wild-derived founder strains CAST/EiJ (CAST) and PWK/PhJ (PWK) were most distinct, underlying much of the genetic variability in protein expression (Extended Data Fig. 1b–d).

We next profiled protein and transcript levels in liver tissue from 192 DO mice, including both females and males, with half of the animals fed standard rodent chow and the other half fed a high-fat diet (Methods, Fig. 1b and Supplementary Tables 2 and 3). In total, we measured 6,756 proteins and 16,921 transcripts with detection in at least half of the samples. Both transcript and protein abundance were highly variable, and principal component analysis identified sex and diet as major drivers of this variation (Extended Data Fig. 2a). As expected, many proteins displayed sex- or diet-specific protein expression. Known female- and male-specific proteins were selectively expressed in a sex-dependent manner (Extended Data Fig. 2b, c). Likewise, many proteins showed diet-specific expression such as PPAR signalling, fat and cholesterol metabolism enzymes (Extended Data Fig. 2d, e), and many of these had concordant transcriptional

<sup>1</sup>Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>The Jackson Laboratory, Bar Harbor, Maine 04609, USA.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.



**Figure 1 | Tandem mass tag (TMT)-based liver proteomics in 192 DO mice.** **a**, Overview of the breeding scheme to create the DO and CC mouse strains. **b**, Experimental overview of the genotyping, transcriptomics and proteomic analysis on 192 DO mouse livers from both sexes on a high-fat or chow diet.

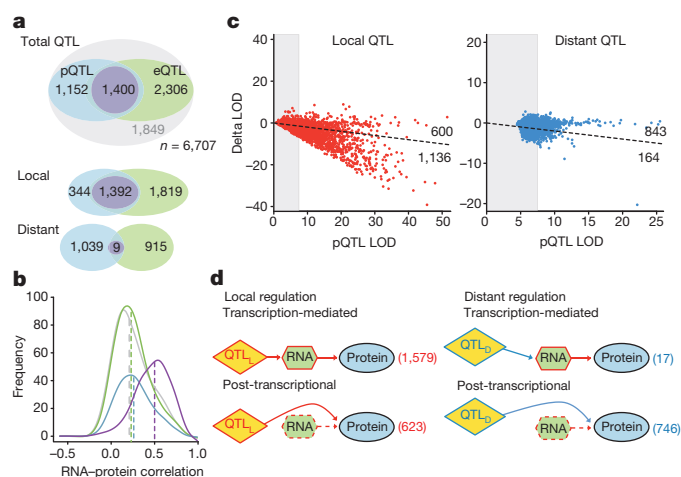
responses (Extended Data Fig. 2f–j). These results demonstrate that sex and diet induced expected changes in transcript and protein expression.

### Genetic regulation of protein abundance

In the subsequent analyses, we focused on 6,707 proteins for which both the protein and its corresponding transcript were detected in at least half of the DO liver samples. Genetic factors explained a substantial portion of variation in the abundance of protein and transcripts in the DO population (Extended Data Fig. 3a–f). To identify these, we performed QTL mapping analysis on transcript (eQTL, Supplementary Table 4) and protein (pQTL, Supplementary Table 5) abundance.

We identified 2,866 pQTL for 2,552 distinct proteins at a genome-wide significance level of  $P < 0.1$  (Fig. 2a). This is the largest set of pQTL identified so far, with tenfold greater numbers than other mass spectrometry (MS)-based approaches. Significant local pQTL were more common than distant pQTL (1,736 local and 1,130 distant pQTL) (Extended Data Fig. 3g). In addition, we identified 4,188 significant eQTL among 3,706 genes, with threefold more local than distant associations at the transcript level (3,211 local and 977 distant eQTL; Fig. 2a, Extended Data Fig. 3h, i). Finally, to examine the replication rate, we analysed a replication set of 192 separate DO mice treated under identical conditions for eQTL (see Methods and Extended Data Fig. 4).

To determine whether the same genetic loci acted on transcript and protein abundance, we first compared the QTL maps. We observed a significant overlap of proteins with pQTL and eQTL ( $n = 1,400$ ; hypergeometric  $P < 1 \times 10^{-16}$ ; Fig. 2a). As expected, genes with concordant QTL had generally higher correlations between protein and transcript abundance compared to those having only pQTL, only eQTL or neither (Fig. 2b). Among local QTL only, we observed a high degree of overlap with 80% of local pQTL having a corresponding local eQTL.

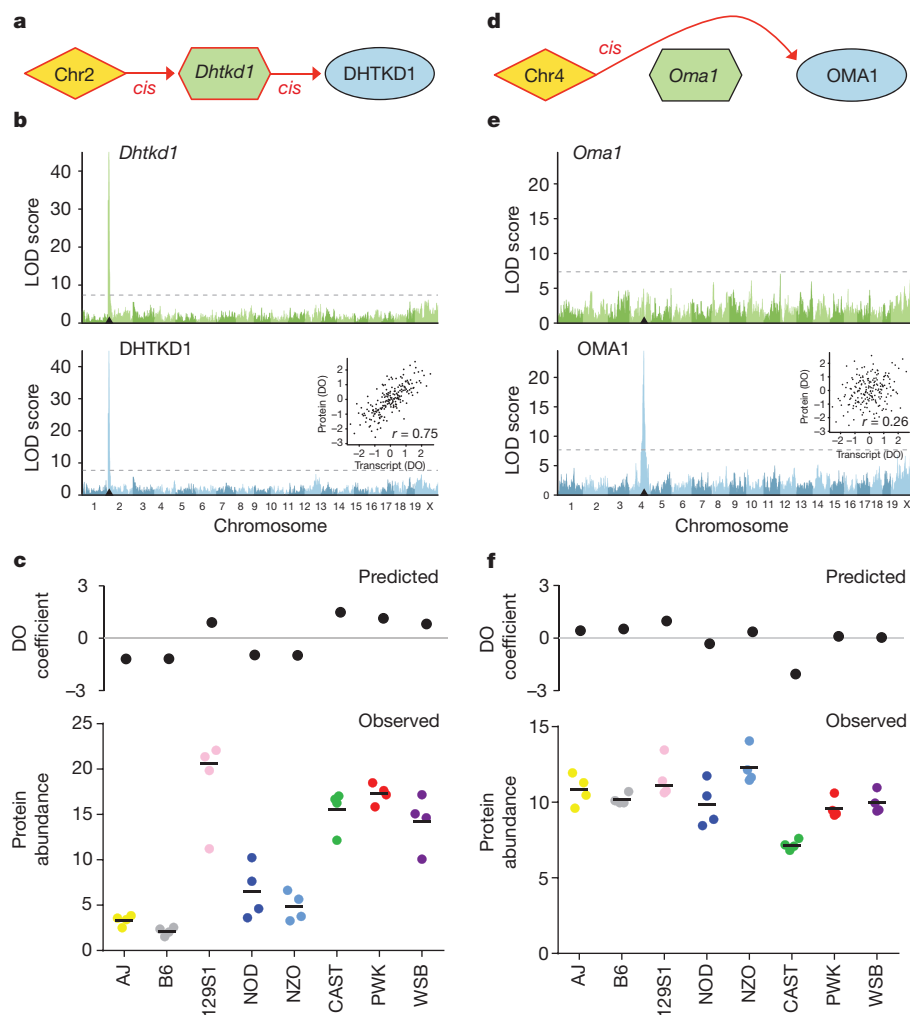


**Figure 2 | Global view of the liver proteome reveals distinct genetic models of protein regulation.** **a**, Venn diagram showing the distribution of transcripts and proteins broken down into local or distant QTL. **b**, Histograms of Pearson correlations for each gene's protein and transcript measurements after segregating into four groups (eQTL–pQTL (purple), pQTL–no eQTL (blue), eQTL–no pQTL (green) and no QTL (grey)). **c**, Local and distant pQTL LOD scores after transcript measurements were used as a covariate in the regression model showing that local pQTL were mediated through their cognate transcripts unlike distant pQTL. **d**, Model selection by Bayesian information criterion (BIC). Local pQTL (QTL<sub>L</sub>) were mostly transcriptionally controlled, whereas distant pQTL (QTL<sub>D</sub>) were regulated generally by post-transcriptional mechanisms.

The small number of local pQTL that lack corresponding eQTL ( $n = 344$ ) could result from genetic variation that regulated protein abundance via post-transcriptional mechanisms such as coding variation that affected protein stability without altering transcript levels. In contrast, distant genetic variants that affected both transcript and protein levels seem to be nearly mutually exclusive (Fig. 2a). This observation leads to the intriguing hypothesis that most distant pQTL affected the abundance of a target protein via post-transcriptional mechanism(s).

For each of the 6,707 expressed proteins, we chose the most significant local and distant QTL, regardless of whether the log odds ratio (LOD) scores at each locus exceeded the pQTL detection threshold. We regressed out the transcript abundance and examined the effect on the peak LOD scores (Fig. 2c). Proteins with pQTL that are mediated through their corresponding transcript should show a reduced LOD score when transcript abundance is included in the regression model. Most local pQTL had significantly lower LOD scores after conditioning on their corresponding transcript (1,136 out of 1,736 dropped by  $\geq 20\%$ ), while most distant pQTL were unaffected after conditioning on their transcript (164 out of 1,007 dropped by  $\geq 20\%$ ). This suggests that local pQTL were largely mediated through transcriptional mechanisms, whereas distant pQTL were more likely to regulate protein abundance without affecting transcript abundance.

We carried out a model selection analysis using Bayesian information criterion (BIC) to identify the most probable path relating a locus genotype to a protein and its corresponding transcript. We evaluated all 6,707 proteins using the best local and distant markers identified in the pQTL mapping, and recorded the path that best explained the observed expression data (Fig. 2d, Extended Data Fig. 5 and Supplementary Tables 6 and 7). We illustrate these models in Fig. 2d in a more simplified form and present a more complete version of these models in Extended Data Fig. 5. Three of the models had no path connecting the locus to protein abundance. For most proteins, these were the best-fitting models for the local QTL ( $n = 4,505$ ) and for the distant QTL ( $n = 5,944$ ). The remaining models linked the abundance of a protein to either a local QTL or a distant QTL. Among



**Figure 3 | Examples of local pQTL that illustrate different models of regulation.** **a**, DHTKD1 abundance is regulated by a local pQTL that probably acts proximally on transcript abundance. **b**, *Dhtkd1* has a strong local eQTL (green) and local pQTL (blue), which corresponds to high correlation between transcript and protein abundance (inset; abundance data transformed to rank normal scores for comparison). **c**, The predicted founder strain abundance of DHTKD1 in the DO population mirrors the measured abundance of DHTKD1 in the founder strains ( $n = 4$  mice for each founder, 2 male and 2 female, black bars represent median values). **d**, OMA1 follows a mode of regulation in which the pQTL acts directly on protein abundance without affecting transcript levels. **e**, OMA1 protein abundance is controlled by a strong local pQTL without a corresponding local eQTL, leading to low correlation (inset) observed between protein and transcript abundance. **f**, The predicted founder strain expression in the DO population is highly correlated to measured OMA1 abundance in the founder strains ( $n = 4$  mice for each founder, 2 male and 2 female, black bars represent median values).

local QTL, we found that most had effects that were mediated at least partially through the transcript ( $n = 1,579$ ), while a minority affected protein abundance independently of the transcript ( $n = 623$ ). Among distant QTL, a much smaller proportion acted through the transcript ( $n = 17$ ), and most affected protein abundance independently of the transcript ( $n = 746$ ). We conclude that most local pQTL affected both protein and transcript abundance, consistent with a transcriptional mode of regulation. However, distant pQTL affected protein abundance independently of the transcript, consistent with a post-transcriptional mode of regulation.

### Local pQTL effect on protein abundance

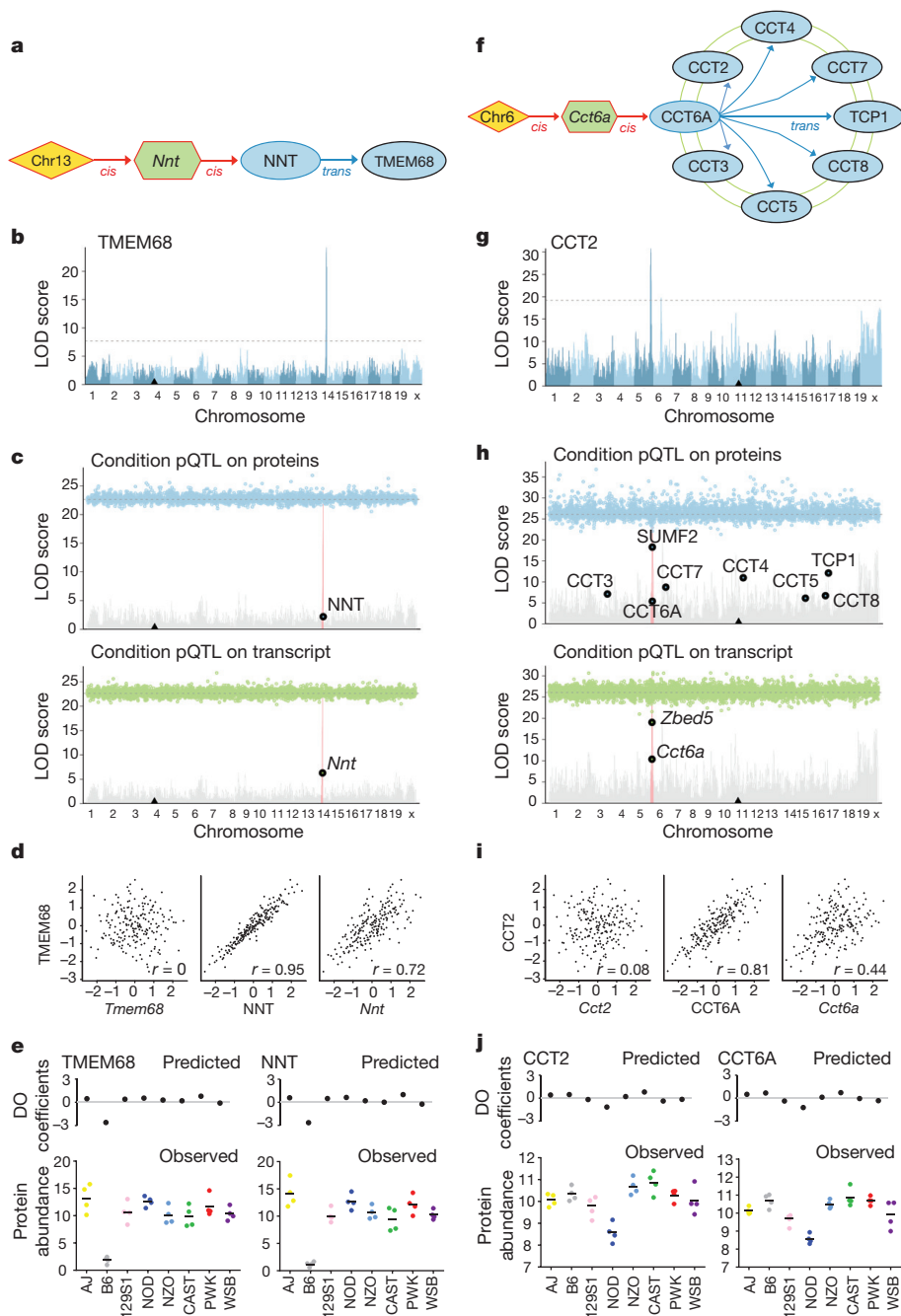
We highlight two examples that illustrate the most common models of local regulation. DHTKD1 exemplifies a pQTL in which a local genetic variant affected transcript abundance that was transmitted to the protein (Fig. 3a, b). This simple transcript-to-protein model of regulation was evidenced by the high correlation between transcript and protein abundance (Fig. 3b, inset) and loss of the pQTL when transcript abundance was added as a covariate in the regression model (Supplementary Table 8). Founder strain allelic contributions derived from the pQTL mapping model suggested that four founder strain alleles (129S1, CAST, PWK and WSB) shared the genetic variant and exhibited higher protein expression levels. To validate these findings, a comparison of these expression coefficients to founder strain data showed the same expression profiles (Fig. 3c). Using genome sequences of the founder strains<sup>26</sup>, we identified a candidate causal genetic variant—a 1-kb deletion in intron 1 of the gene. The same variant was previously reported as a pQTL in the DBA mouse strain<sup>17</sup>. DHTKD1 was just one of almost 1,600 cases in which

QTL-to-transcript-to-protein regulation was identified as the best local model. Additional examples include *Ces2h* and *Pipox* (Extended Data Fig. 6).

A total of 623 proteins had local pQTL that affected protein abundance directly, including OMA1 (Fig. 3d–f). These proteins were uncoupled from their transcript, as evidenced by the lack of correlation between protein and transcript abundance (Fig. 3e, inset). For *Oma1*, founder allele contributions in the DO population pointed to an allele from the CAST strain causing reduced protein levels. This was validated by protein expression in the founder strains (Fig. 3f). Genome analysis identified four missense mutations in *Oma1* (H73N, R97Q, I127K and V283L), suggesting that protein structure may be affected and not the transcript. Other examples of variants that influenced protein expression that were not mediated through transcripts include *Entpd5* and *Lars2* (Extended Data Fig. 6).

### Causal intermediates of distant pQTL

Unlike local pQTL, in which the causative variant is directly linked to the target protein-coding gene, distant pQTL exert their effects on target proteins *in trans* through a causal intermediate. To determine whether a distant pQTL acts proximally through the transcript of the affected protein or directly on the protein bypassing the transcript, we used mediation analysis (see Methods). We examined 1,130 distant pQTL and identified at least one candidate protein or transcript mediator for 743 (Supplementary Table 8). In total, we found 618 unique protein/transcript mediators, of which 534 regulated a single protein, 61 regulated two proteins, and 23 regulated three or more proteins. Furthermore, 84% of the top candidate protein mediators were themselves driven by a local pQTL. This illustrates that a single local QTL,



**Figure 4 | Mediation of distant pQTL reveals network interactions in the liver proteome.**

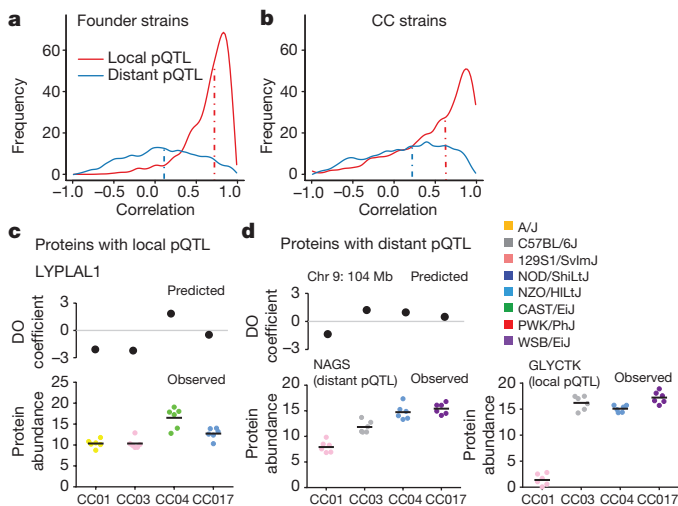
**a**, The genetic variant underlying the distant TMEM68 pQTL acts proximally in *cis* on *Nnt* transcript and protein abundance. **b**, TMEM68 protein abundance is buffered against local genetic variation affecting transcript levels by a distant regulator on chromosome 13. **c**, Mediation analysis identified NNT protein and *Nnt* transcript as the likely mediator. **d**, TMEM68 protein is poorly correlated to its corresponding transcript, but highly correlated with both NNT protein and *Nnt* transcript abundance. **e**, TMEM68 strain abundance predicted at the chromosome 13 distant pQTL in the DO population is highly correlated to TMEM68 and NNT abundance measured in the founder strains, and matches the predicted NNT strain abundance in the DO population ( $n = 4$  mice for each founder, 2 male and 2 female, black bars represent median values). In all cases the C57BL/6J allele is observed to be the low expressor. **f**, The chromosome 5 variant responsible for the distant effect on CCT2 abundance acts proximally in *cis* on *Cct6a* transcript and protein abundance. **g**, All members of the chaperonin containing Tcp1 (CCT) complex including CCT2 exhibit a distant pQTL that maps to distal chromosome 5. **h**, Mediation analysis identified *Cct6a*/CCT6A as the probable mediator of this effect. Protein mediation shows that the protein abundance of all CCT complex members is highly correlated as all members are pulled down in the background of the mediation plot. **i**, CCT2 protein abundance is highly correlated to CCT6A protein and *Cct6a* transcript abundance. All other CCT complex members show this same pattern. **j**, CCT2 abundance predicted at the chromosome 5 distant pQTL is highly correlated with CCT2 and CCT6A abundance measured in the founder strains, and tracks with CCT6A abundance predicted at the pQTL in the DO population ( $n = 4$  mice for each founder, 2 male and 2 female, black bars represent median values). DO animals that derive the chromosome 5 region from NOD/ShiLtJ have lower abundance of all CCT proteins.

acting proximally on a transcript or protein intermediate, can effectively control the abundance of a distant protein or multiple distant proteins, uncoupling them from their transcriptional control mechanisms.

We highlight examples in which mediation analysis identified the regulatory protein or transcript underlying the distant pQTL. TMEM68 protein exemplified a post-transcriptional model of regulation (Fig. 4a). TMEM68 has a distant pQTL peak on chromosome 13, and the *Temem68* transcript has a local QTL on chromosome 4 (Fig. 4b, Supplementary Table 4). The protein and transcript levels were uncorrelated (Fig. 4d, left). We identified both NNT protein and *Nnt* transcript on chromosome 13 as candidate mediators of the distant pQTL for TMEM68 (Fig. 4c). The *Nnt* protein and transcript shared a local QTL indicating a transcriptional mechanism. Both *Nnt* protein and transcript were highly correlated with TMEM68 abundance (Fig. 4d). Founder allele expression patterns inferred at the distant pQTL suggest that a variant in B6 mice causes a downregulation in NNT protein levels, which was validated by proteomic analysis of the founder

strains (Fig. 4e). This effect on *Nnt* expression has been previously attributed to a small exonic deletion found only in the B6 strain<sup>28–30</sup>. Using this same approach, we reconfirmed numerous known protein–protein associations including SNX7–SNX4, PGAM1–PGAM2 and LRRFIP1–FLII (refs 31–33), and inferred many new associations (Extended Data Fig. 7).

The chaperonin containing TCP1 (CCT) complex illustrates how mediation analysis can reveal larger co-regulated complexes and pathways (Fig. 4f). All eight subunits of the CCT complex shared a distant pQTL (but not distant eQTL) on chromosome 5 with the same pattern of allele effects. We identified the transcript and protein abundance of *Cct6a* as mediators of this post-transcriptional distant effect (Fig. 4g, h). This relationship is evident by the high correlation in protein–protein and protein–transcript abundance between *Cct6a* and other complex members (Fig. 4i, Extended Data Fig. 8). Founder strain allele effects inferred at the distant pQTL showed that DO animals containing the NOD strain allele on chromosome 5 expressed lower



**Figure 5 | Genotype can be an accurate predictor of protein abundance.**

**a**, Founder strain protein abundance values inferred at significant pQTL in the DO population closely match measured abundance values from the founder strains themselves. The distributions of Pearson correlations are plotted for local pQTL and distant pQTL. Local pQTL are generally more predictive of abundance values in the founder strains (local median  $r = 0.72$ , distant median  $r = 0.11$ ). **b**, Founder strain allele predictions from the DO were also assessed against protein abundance data collected from four CC strains ( $n = 6$  mice per strain). We observe that local pQTL are more predictive of protein abundance in the CC strains (local median  $r = 0.63$ ; distant median  $r = 0.22$ ). **c**, Predictive power depends largely on the significance of the pQTL. Local pQTL generally had higher LOD scores, and as such we had higher power to predict these proteins ( $n = 4$  mice for each founder, 2 male and 2 female, black bars represent median values). An example is shown for LYPLAL1. **d**, Protein abundance could also be predicted for genes with significant distant pQTL in the DO population; however, as a group these predictions were modest compared to local pQTL. As an example, NAGS abundance in the CC strains could be predicted based on the local genotype at its mediator protein, GLYCTK ( $n = 6$  mice for each CC strain, 3 male and 3 female, black bars represent median values).

overall levels of the entire complex. This same pattern was observed in the founder strains (Fig. 4j). Genome sequence analysis identified a variant (rs228180583) in a conserved KLF4-binding domain in the *Ctcf* promoter region that was present only in the NOD strain. From these data, we propose that the variant lowers *Ctcf* transcript and protein abundance, which results in a stoichiometric imbalance and degradation of excess unbound complex members. These examples highlight the power of mediation analysis to identify protein–protein associations and co-regulated groups of proteins.

### Genetic perturbations reveal protein networks

By leveraging the large number of distant pQTL and mediation analysis of each, we created a network of pQTL-regulated proteins (Extended Data Fig. 9a). Each distant pQTL is causally linked to its target protein with mediators and other co-regulated proteins to form a network. When merged across all 1,130 distant pQTL, the network comprises 5,794 causal or co-regulatory relationships among 3,938 proteins or QTL. Markov cluster algorithm (MCL) clustering defined 671 clusters of variable sizes (Extended Data Fig. 9b). Approximately 44% of clusters included members with shared biological functions as assessed by Gene Ontology (GO) enrichment (Extended Data Fig. 9c). As an example, almost all cholesterol synthesis enzymes were determined to be co-regulated and associated with just two distant pQTL that affected the protein expression for *Lss* and *Cyp51* (Supplementary Table 8). Clusters found within the larger regulatory network tended to associate proteins with shared biological properties. Some clusters grouped proteins according to subcellular localization, as seen for complex I

of the electron transport chain (Extended Data Fig. 9d), SUCLG1/SUCLG2 and associated mitochondrial proteins (Extended Data Fig. 9e), and IMMT/SAMM50 with other mitochondrial proteins (Extended Data Fig. 9f). Each corresponds to a well-studied complex, suggesting that the regulatory network emerging from mediation analysis provides an accurate snapshot of mouse liver gene regulation.

To probe further the correspondence between protein co-regulation and physical association, each pQTL and its co-regulated proteins were mapped onto an ongoing and recently published human interactome network<sup>34</sup>. Physical associations accounted for a significant subset of protein regulatory networks, especially among distant QTL (Extended Data Fig. 9g–i). Through these findings, we propose that a considerable fraction of distant pQTL were the direct result of post-transcriptional regulation of proteins that had similar biological functions, cell locations, and/or complex membership.

### Genotype is a predictor of protein abundance

For many genes with pQTL, founder strain allele patterns inferred from the DO pQTL mapping model closely matched protein abundance measured in the founder strains themselves. To determine the extent to which genotype can be a predictor of protein abundance, we examined all significant pQTL and compared the founder strain coefficients observed at the pQTL location to the protein levels measured in the founder strains (Fig. 5a). We found that predictive power increased with the significance of the pQTL (Fig. 5a, Extended Data Fig. 10). Because of their tight linkage to the controlled gene, local pQTL tended to have higher predictive power than distant loci (local pQTL median  $r = 0.72$ ; distant pQTL median  $r = 0.11$ ). However, highly significant distant pQTL ( $>10$  LOD) have comparable predictive power to local pQTL of similar significance.

We further validated our strains predictions using the quantitation of ~6,500 proteins from four CC strains (Supplementary Table 9). For each pQTL, we identified the genotype in the CC strains and predicted the protein abundance using the DO proteomics data. Our data suggest that strain genotype is also predictive of protein abundance in the CC strains (Fig. 5b). The predictive power was higher for local pQTL than distant ones. As an example, LYPLAL1 was identified with a local pQTL in the DO population and was predicted to have lower protein abundance in the CC001 and CC003 strains (Fig. 5c). For distant pQTL with high LOD scores, the predictive power was also high. For distant pQTL, these predictions were made by comparing the measured protein and the genotype at the QTL location. For example, GLYCTK protein abundance was predicted using the genotype at the *Nags* gene location where the variant was detected (Fig. 5d).

This study quantified both protein and transcript abundance in a genetically diverse population of mice, mapping their genetic architecture. We identified the largest catalogue of pQTL so far, which can be attributed to two variables in our experimental design. First, we have improved the accuracy and sensitivity of quantification for both protein and transcript abundance. Second, our experimental population captured genetic diversity far in excess of the human population and standard laboratory mouse strains. Earlier studies reported a disconnect between transcript and protein abundance<sup>2,3,6</sup>, which has also been a conclusion drawn from several recent eQTL–pQTL analyses<sup>4,7,17,35</sup>. Data here show that local QTL tend to abide by the central dogma as demonstrated by concordant effects on transcripts and proteins, whereas distant pQTL are conferred by post-transcriptional mechanisms. Our mediation analysis provided the ability to identify causal protein intermediates underlying distant pQTL and led to the identification of hundreds of protein–protein associations. Our experimental design provides an advantage over protein interaction maps because genetic mapping is not dependent on physical interactions. This conclusion is further exemplified by the co-regulation of protein complexes or biochemical pathways in this study. Stoichiometric buffering provides one explanation for co-regulation of protein complexes and may account for earlier observations that protein abundances

(but not transcript abundances) of orthologues are well-conserved across large evolutionary distances<sup>36,37</sup>.

These findings suggest a new predictive genomics framework in which quantitative proteomics and transcriptomics are combined in the analysis of a discovery population like the DO to identify genetic interactions. Next, pathways relevant to the tissue/physiological phenotype of interest are intersected with the list of significant pQTL. Pathways enriched for proteins with significant pQTL should be amenable to manipulation in the founder and CC strains. That is, the founder allele effects inferred at the pQTL can be combined in such a way via crosses of CC strains to tune pathway output. Moreover, as we better understand the types of mutation that can affect protein abundance, we can introduce specific mutations with gene editing into sensitized or robust genetic backgrounds. We foresee this strategy being used to design reproducible rodent models that span a range of human-relevant phenotypes, for example, in drug metabolism or toxicology studies.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 22 September 2015; accepted 13 April 2016.**

**Published online 15 June 2016.**

- Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
- Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730 (1999).
- Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
- Ghazalpour, A. *et al.* Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet.* **7**, e1001393 (2011).
- Skelly, D. A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res.* **23**, 1496–1504 (2013).
- Wühr, M. *et al.* Deep proteomics of the *Xenopus laevis* egg using an mRNA-derived reference database. *Curr. Biol.* **24**, 1467–1475 (2014).
- Fu, J. *et al.* System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nat. Genet.* **41**, 166–167 (2009).
- Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **7**, 862–872 (2006).
- Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
- Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
- Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
- Jansen, R. C. & Nap, J. P. Genetical genomics: the added value from segregation. *Trends Genet.* **17**, 388–391 (2001).
- Chesler, E. J. *et al.* Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.* **37**, 233–242 (2005).
- Foss, E. J. *et al.* Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLoS Biol.* **9**, e1001144 (2011).
- Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nat. Genet.* **39**, 1369–1375 (2007).
- Khan, Z., Bloom, J. S., Garcia, B. A., Singh, M. & Kruglyak, L. Protein quantification across hundreds of experimental conditions. *Proc. Natl Acad. Sci. USA* **106**, 15544–15548 (2009).
- Wu, Y. *et al.* Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. *Cell* **158**, 1415–1430 (2014).
- Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499**, 79–82 (2013).
- Damerval, C., Maurice, A., Josse, J. M. & de Vienne, D. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* **137**, 289–301 (1994).
- Albert, F. W., Treusch, S., Shockley, A. H., Bloom, J. S. & Kruglyak, L. Genetics of single-cell protein abundance variation in large yeast populations. *Nature* **506**, 494–497 (2014).
- Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **8**, 937–940 (2011).
- McAlister, G. C. *et al.* MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
- Churchill, G. A. *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* **36**, 1133–1137 (2004).
- Churchill, G. A., Gatti, D. M., Munger, S. C. & Svenson, K. L. The Diversity Outbred mouse population. *Mamm. Genome* **23**, 713–718 (2012).
- Threadgill, D. W. & Churchill, G. A. Ten years of the Collaborative Cross. *Genetics* **190**, 291–294 (2012).
- Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
- Gatti, D. M. *et al.* Quantitative trait locus mapping methods for Diversity Outbred mice. *G3 (Bethesda)* **4**, 1623–1633 (2014).
- Toye, A. A. *et al.* A genetic and physiological study of impaired glucose homeostasis control in C57BL/6J mice. *Diabetologia* **48**, 675–686 (2005).
- Ronchi, J. A. *et al.* A spontaneous mutation in the nicotinamide nucleotide transhydrogenase gene of C57BL/6J mice results in mitochondrial redox abnormalities. *Free Radic. Biol. Med.* **63**, 446–456 (2013).
- Freeman, H. C., Hugill, A., Dear, N. T., Ashcroft, F. M. & Cox, R. D. Deletion of nicotinamide nucleotide transhydrogenase: a new quantitative trait locus accounting for glucose intolerance in C57BL/6J mice. *Diabetes* **55**, 2153–2156 (2006).
- Huttlin, E. L. *et al.* The BioPlex Network: a systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015).
- van Weering, J. R. T. *et al.* Molecular basis for SNX-BAR-mediated assembly of distinct endosomal sorting tubules. *EMBO J.* **31**, 4466–4480 (2012).
- Liu, Y.-T. & Yin, H. L. Identification of the binding partners for flightless I, A novel protein bridging the leucine-rich repeat and the gelsolin superfamily. *J. Biol. Chem.* **273**, 7920–7927 (1998).
- Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
- Battle, A. *et al.* Genomic variation. Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667 (2013).
- Laurent, J. M. *et al.* Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics* **10**, 4209–4212 (2010).
- Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
- Welsh, C. E. *et al.* Status and access to the Collaborative Cross population. *Mamm. Genome* **23**, 706–712 (2012).
- Chesler, E. J. *et al.* The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. *Mamm. Genome* **19**, 382–389 (2008).
- Iraqi, F. A., Churchill, G. & Mott, R. The Collaborative Cross, developing a resource for mammalian systems genetics: a status report of the Wellcome Trust cohort. *Mamm. Genome* **19**, 379–381 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors thank L. Somes and M. Strobel for breeding the mice; S. Ciciotte, S. Daigle, J. Pereira, C. Snow, R. Lynch and H. Munger for extracting RNA and performing RNA-seq experiments; and A. Manichaik for discussion on mediation analysis. Collaborative Cross strains used in this study were imported to JAX from the Systems Genetics Core Facility at the University of North Carolina (USA)<sup>38</sup>. Previous to their relocation to UNC, CC lines CC001, CC003 and CC017 were generated and bred at Oak Ridge National Laboratory (USA)<sup>39</sup>; CC line CC004 was generated and bred at Tel Aviv University (Israel)<sup>40</sup>. Research reported here was supported by Harvard Medical School, The Jackson Laboratory, and National Institutes of Health (NIH) grants under awards P50GM076468 (to G.A.C.), F32HD074299 (to S.C.M.), GM67945 (to S.P.G.) and U41HG006673 (to S.P.G. and E.L.H.).

**Author Contributions** S.P.G. and G.A.C. conceived and planned the study. J.M.C. performed the proteomics sample preparation and mass spectrometry analysis. S.C.M., K.C., N.R., P.S. and D.M.G. performed the RNA-seq, genetic mapping, model selection and mediation analyses. S.C.M. and G.A.C. conceived the mediation analysis. E.L.H. performed the network analysis for distant pQTL. K.L.S. and G.A.C. provided DO and CC mice. J.M.C., S.C.M., G.A.C. and S.P.G. interpreted the scientific findings, discussed and wrote the paper with input from all authors.

**Author Information** The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD002801 (<http://www.proteomexchange.org/>). Raw RNA-seq fastq files and processed gene-level data are archived at Gene Expression Omnibus (GEO) under accession number GSE72759. We implemented our mediation method as the R package, intermediate, which can be freely downloaded from <http://github.com/churchill-lab/intermediate>. The Genotyping by RNA-seq (GBRS) software is available for download from <https://github.com/churchill-lab/gbrs>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.P.G. ([steven\\_gygi@hms.harvard.edu](mailto:steven_gygi@hms.harvard.edu)) or G.A.C. ([gary.churchill@jax.org](mailto:gary.churchill@jax.org)).

## METHODS

The sample size (192 animals) was calculated based on previous experimental RNA-seq data and was determined to be sufficient to detect genetics effects that explain 10% or more genetic variation with 90% power and  $10^{-6}$  type I error rate. Randomization was used to assign mice to treatments and samples to batches, bar codes, and TMT tags in both the RNA-seq and proteomics experiments. Data collection was carried out by automation, and as such there was no need for blinding the sample identifiers.

**Animals and genotyping: DO mice.** Diversity Outbred mice (DO, stock no. 009376) were obtained from The Jackson Laboratory (JAX) at 3 weeks of age, housed at JAX, and fed either standard rodent chow (6% fat by weight, LabDiet 5K52; LabDiet, Scott Distributing) or a high-fat diet (44.6% kcal fat and 34% kcal sucrose by weight, TD.08811, Harlan Laboratories) from wean age throughout the study. In total, 192 DO mice were analysed in the current study, including 50 females and 48 males raised on standard chow, and 48 females and 46 males raised on the high fat diet. At 26 weeks of age, animals were euthanized, dissected, and liver samples were sent for RNA-seq analysis at JAX (samples stored in RNAlater solution; Life Technologies) and proteomics analysis at Harvard Medical School (HMS; samples sent as snap frozen tissue).

**Animals and genotyping: founder strain and CC mice.** Two male and two female mice from each of the eight DO/CC founder inbred strains and four (3 males and 3 females) CC recombinant inbred strains (CC strains CC001, CC003, CC004, and CC017) were obtained from and housed at JAX, raised on the standard chow diet. Founder strain mice were euthanized at 26 weeks of age, and the CC mice were euthanized at 8–16 weeks of age. Liver samples were dissected from each mouse, snap frozen and sent to HMS for proteomics analysis. All procedures on mice were approved by the Animal Care and Use Committee at JAX.

**Multiplexed quantitative proteomic analysis of mouse livers: sample preparation and TMT labelling.** A total of 192 DO mouse livers (~50 mg), 32 founder strains livers (8 founders strains, 2 male and 2 female replicates for each strain) and 24 CC strain livers (4 strains, 3 male and 3 female replicates for each strain) were homogenized in 1 ml lysis buffer (1% SDS, 50 mM Tris, pH 8.8 and Roche complete protease inhibitors). Samples were reduced with 5 mM dithiothreitol for 30 min at 37 °C followed by alkylation with 15 mM for 30 min at room temperature in the dark. The alkylation reaction was quenched by adding 5 mM dithiothreitol for 15 min at room temperature in the dark. A 500 µl aliquot was then methanol/chloroform precipitated. The samples were allowed to air dry before being resuspended in 1 ml of 8 M urea and 50 mM Tris, pH 8.8. The urea concentration was diluted down to ~1.5 M urea with 50 mM Tris. Proteins were quantified using a BCA assay. Protein was then digested using a combination of Lys-C/trypsin at an enzyme-to-protein ratio of 1:100. First, protein was digested overnight with Lys-C followed by 6-h digestion with trypsin all at 37 °C. Samples were then acidified using formic acid to approximately pH 3. Samples were then desalted using a SepPak column. Eluents were then dried using a vacuum centrifuge. Peptide pellets were resuspended in 110 µl of 200 mM HEPES buffer, pH 8, and peptides were quantified by a BCA assay. Approximately 70 µg of peptides (100 µl of sample + 30 µl of 100% acetonitrile) were then labelled with 15 µl of 20 µg µl<sup>-1</sup> of the corresponding TMT 10-plex reagent (DO or founder strains) or TMT 8-plex reagent (CC strains) for 2 h at room temperature. The reaction was quenched using 8 µl of 5% hydroxylamine for 15 min. Peptides were then acidified using 150 µl of 1% formic acid, each set of 10 samples were mixed and desalted using a SepPak column. In total, 25 TMT 10-plex reactions and 3 8-plex reactions were performed (21 DO mice, 4 founder strains and 3 CC strains). The full labelling schemes for the DO mice, the founder strains and CC strains are provided as supplementary tables (Supplementary Tables 1, 3 and 7).

**Basic reverse-phase fractionation.** Each of the 28 TMT experiments was separated by basic, reversed-phase chromatography. Samples were loaded onto an Agilent 300 Extend C18 column (5 µm particles, 4.6 mm ID and 220 mm in length). Using an Agilent 1100 quaternary pump equipped with a degasser and a photodiode array detector (set at 220- and 280-nm wavelength), peptides were separated using a 50 min linear gradient from 18% to 40% acetonitrile in 10 mM ammonium bicarbonate, pH 8, at a flow rate of 0.8 ml min<sup>-1</sup>. Peptides were separated into a total of 96 fractions that were consolidated into 24. Samples were subsequently acidified with 1% formic acid and vacuum centrifuged to near dryness. Each fraction was desalted via StageTip, dried via vacuum centrifugation, and reconstituted in 1% formic acid for liquid chromatography tandem mass spectrometry (LC–MS/MS) processing.

**Liquid chromatography electrospray ionization tandem mass spectrometry (LC–ESI–MS/MS).** Peptides from every odd fraction (12 fractions total) from basic reverse-phase fractionation were analysed using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) equipped with a Proxeon ultra high pressure liquid chromatography unit. Peptide mixtures were separated on a 100 µm

ID microcapillary column packed first with ~0.5 cm of 5 µm Magic C18 resin followed by 40 cm of 1.8 µm GP-C18 resin. Peptides were separated using a 3-h gradient of 6–30% acetonitrile gradient in 0.125% formic acid with a flow rate of ~400 nl min<sup>-1</sup>. In each data collection cycle, one full MS scan (400–1,400 *m/z*) was acquired in the Orbitrap (1.2 × 10<sup>5</sup> resolution setting and an automatic gain control (AGC) setting of 2 × 10<sup>5</sup>). The subsequent MS2–MS3 analysis was conducted with a top 10 setting or a top speed approach using a 2-s duration. The most abundant ions were selected for fragmentation by collision induced dissociation (CID). CID was performed with a collision energy of 35%, an AGC setting of 4 × 10<sup>3</sup>, an isolation window of 0.5 Da, a maximum ion accumulation time of 150 ms and the rapid ion trap setting. Previously analysed precursor ions were dynamically excluded for 40 s.

During the MS3 analyses for TMT quantification, precursors were isolated using a 2.5–Da *m/z* window and fragmented by 35% CID in the ion trap. Multiple fragment ions (SPS ions) were co-selected and further fragmented by HCD. Precursor ion selection was based on the previous MS2 scan and the MS2–MS3 was conducted using sequential precursor selection (SPS) methodology. HCD used for the MS3 was performed using 55% collision energy and reporter ions were detected using the Orbitrap with a resolution setting of 60,000, an AGC setting of 50,000 and a maximum ion accumulation time of 150 ms.

**Database searching and reporter ion quantification.** Software tools were used to convert mass spectrometric data from raw file to the mzxml format<sup>34</sup>. Erroneous charge state and monoisotopic *m/z* values were corrected as per previous publication<sup>34</sup>. MS/MS spectra assignments were made with the Sequest algorithm<sup>41</sup> using an indexed Ensembl database (mouse: *Mus musculus* NCBIM37.61). Databases were prepared with forward and reversed sequences concatenated according to the target-decoy strategy<sup>42</sup>. All searches were performed using a static modification for cysteine alkylation (57.0215 Da) and TMT on the peptide N-termini and lysines. Methionine oxidation (15.9949 Da) was considered a dynamic modification. Mass spectra were searched with trypsin specificity using a precursor ion tolerance of 10 p.p.m. and a fragment ion tolerance of 0.8 Da. Sequest matches were filtered by linear discriminant analysis as described previously, first to a data set level error of 1% at the peptide level based on matches to reversed sequences<sup>42</sup>. Peptide probabilities were then multiplied to create protein rankings and the data set was again filtered to a final data set level error of 1% false discovery rate (FDR) at the protein level. The final peptide-level FDR fell well below 1% (~0.2% peptide level). Peptides were then assigned to protein matches using a reductionist model, where all peptides were explained using the least number of proteins.

Peptide quantitation using TMT reporter ions was accomplished as previously published<sup>21,22</sup>. In brief, a 0.003 Da *m/z* window centred on the theoretical *m/z* value of each reporter ion was monitored for each of the 8–10 reporter ions, and the intensity of the signal closest to the theoretical *m/z* value was recorded. TMT signals were also corrected for isotope impurities based on the manufacturer's instructions. Peptides were only considered quantifiable if the total signal-to-noise for all channels was >200 and an isolation specificity of >0.75. Within each TMT experiment, peptide quantitation was normalized by summing the values across each channel and then each channel was corrected so that each channel had the same summed value. Protein quantitation was performed by summing the signal-to-noise for all peptides for a given protein. Protein quantitative measurements were then scaled to 100 (equal expression across all channels would be a value of 10). Normalization across each of the 10plex experiments was then performed using quantile normalization.

**Statistical analyses.** Principal component analysis was performed using Cluster 3.0 (ref. 43). Hierarchical clustering, *K*-means clustering and ANOVA were performed using Multiexperiment Viewer. Analysis on the founder strains proteomics data sets was performed using an ANOVA and adjusted for multiple testing using the Benjamini–Hochberg FDR procedure.

**Implications of multiplexed quantitative proteomics platform.** Improvements in several aspects of the analysis pipeline enabled the increase in scale. Our quantitative proteomics technology proved instrumental as it supported multiplexing with ten different mouse livers in the same analysis. Accurate expression measurements were obtained by applying a notched isolation waveform on an Orbitrap Fusion instrument. The time required to collect expression profiles from each 10-plex was 36 h or ~4 h per mouse liver of mass spectrometry analysis time. The proteome-wide analysis of 192 livers thus required 35 days. As a result of these methodology improvements, we detected tenfold more pQTL than previous MS-based reports.

**Genotyping of DO and CC samples: DO samples.** Genomic DNA was extracted from each DO mouse (*n* = 192 total samples) and genotyped at 57,973 single nucleotide polymorphisms (SNPs) on the Mega-MUGA platform (Geneseek)<sup>44</sup>. A total of 177 out of 192 samples passed SNP quality control metrics. For these samples, founder haplotypes were inferred from SNP probe intensities using a

hidden Markov model implemented in the DOQTL R package<sup>27,45</sup>, and then used to interpolate a grid of 64,000 evenly-spaced genetic intervals. In addition, founder haplotypes were independently inferred from the RNA-seq data by genotyping by RNA-seq (GBRS) protocol (see next section) and interpolated to the same 64,000 interval grid.

For each sample, we verified that the haplotype reconstructions agreed between the DNA Mega-MUGA and GBRS reconstructions by calculating the Pearson correlation between each pair of samples. When a Mega-MUGA sample had a correlation below 0.4 with the same sample ID in the RNA-seq data, we assumed that this sample was mismatched. We searched the RNA-seq data for the correct match to the Mega-MUGA sample by looking for another sample that was more highly correlated. If we found an RNA-seq sample with a correlation >0.4 that was not assigned to another sample, we matched it with the Mega-MUGA sample. When a sample was removed from the Mega-MUGA data for technical reasons, we used the GBRS haplotype reconstructions (samples F326, F328, F362, F363, F368, M377, M388, M392, M393, M394, M404, M408, M411, M419 and M425). **Genotyping of DO and CC samples: CC samples.** Founder haplotypes for the CC strains were downloaded from the CC strain database (csbio.unc.edu/CCstatus/gstemp/AllImageHapAndGenotypeFiles.zip) maintained at the University of North Carolina.

**Transcriptome profiling and GBRS.** Total liver RNA was isolated from each of the 192 DO mice and sequenced by single-end RNA-seq as previously described<sup>46</sup>. We aligned raw reads against pooled transcriptomes of the eight founder strains. To construct the pooled transcriptome, we incorporated founder strain-specific SNPs and insertions/deletions (Sanger REL-1410) into the reference strain genome sequence (GRCm38/mm10) to produce strain-specific genomes. We derived transcript sequences for all annotated genes (Ensembl version 75 gene annotation) from each strain genome, and then combined the eight founder allele sequences for each transcript into one pooled transcriptome for read alignment. After alignment, we quantified expected read counts expressed from each transcript allele using an expectation maximization algorithm (EMASE, <https://github.com/churchill-lab/emase>). We repeated the same process for liver RNA-seq data from the eight founder strains to assess how specifically each founder read aligns back to their origin strain when exposed to all other founder alleles simultaneously in the alignment pool. We then evaluated the genotype probability of each transcript using a hidden Markov Model (HMM), where we bring those read counts together and calculate (1) how likely allele-specific read counts are generated from a specific genotype, and (2) how much those likelihoods comply within the context of neighbouring transcripts. Finally, we re-quantified total and allele-specific expression with EMASE by repeating the similar process but using individualized diploid transcriptomes reconstructed along our genotype calls.

**QTL mapping of transcript and protein abundance.** Quantitative proteomics combined with transcript quantitation by RNA-seq makes it possible to define the relative contributions of transcriptional versus post-transcriptional mechanisms and local versus distant effects on protein abundance. For example, a local QTL is a genetic variant near the target gene that influences its expression; it might be expected to act in *cis* and affect both transcript and protein levels. By contrast, distant QTL exert their effect on a target gene's expression in *trans*, most likely via a causal intermediate such as another protein or RNA species. Identifying causal intermediates of distant QTL effects may reveal novel protein–protein associations and their biological consequences. Our comprehensive pQTL analysis yielded a global network of interactions that shed new light on the regulation of protein abundance.

**QTL mapping.** For mapping of pQTL and eQTL, we included only proteins that were present (non-0) in  $\geq 96$  samples and corresponded to gene identifiers in the RNA-seq data that were also expressed in  $\geq 96$  samples. A total of 6,707 proteins met these criteria. For pQTL mapping with the proteomics data, protein abundance values were first quantile-normalized and transformed to rank normal scores, and then pQTL were mapped with the R package DOQTL<sup>27</sup>, using a linear mixed model with sex, diet and TMT tag as additive covariates and a random polygenic term to account for genetic relatedness among the DO animals<sup>47</sup>. For eQTL mapping from the RNA-seq data, gene-level counts were first normalized to the upper quartile value and transformed to rank normal scores, and then eQTL were mapped with DOQTL including sex, diet and batch as additive covariates and a random polygenic term to account for relatedness. We used the 64 k genotype matrix derived from Mega-MUGA DNA genotypes as input for pQTL and eQTL mapping, with the exception of samples with missing or low quality DNA genotype results where we used GBRS-derived genotypes.

**Statistical analyses.** Significance thresholds were established by performing 10,000 permutations and fitting an extreme value distribution to the maximum LOD scores<sup>48</sup>. Permutation derived *P* values were then converted to *q*-values with the QVALUE R package, using the bootstrap method to estimate  $\pi_0$  and the default  $\lambda$

tuning parameters<sup>49</sup>. The significance threshold for declaring a QTL was set at a genome-wide significance level of  $P < 0.1$  (FDR = 10%).

**eQTL replication analysis.** To detect a pQTL and eQTL requires a strong statistical signal to exceed stringent genome-wide significance thresholds. We considered the possibility that lack of concordance between distant pQTL and eQTL could be explained by low power, especially for the distant pQTL. The proteomics data in this study were obtained on a subset (discovery set) of DO mice from an earlier study<sup>46</sup>. We created a replication set for the eQTL by random sampling of 192 additional DO samples. As expected, the likelihood of replicating an eQTL depended on the significance of the QTL in the discovery set (Supplementary Fig. 4a). Local eQTL tend to be more significant and replicated well across experiments (76% replication,  $n = 2,448$ ), while distant eQTL replicated poorly (5% replication,  $n = 52$ ; Supplementary Fig. 4a). The distribution of LOD scores is similar for distant pQTL and distant eQTL (Supplementary Fig. 4b), suggesting that we had similar low power to detect distant pQTL as distant eQTL. While the overlap between distant pQTL and eQTL is lower than what we had expected (<1%,  $n = 9$ ; Supplementary Fig. 4c), it is still difficult to rule out low rate of detection as a possible explanation. We provide additional evidence that distant pQTL act through post-transcriptional mechanisms.

**Model selection by BIC.** For each of the 6,707 proteins in the discovery set with detectable transcript and protein abundance, we identified (1) the locus within  $\pm 10$  Mb of the gene midpoint with the highest LOD score (local), and (2) the locus on a separate chromosome with the highest LOD score (distant), regardless of their statistical significance. Next, for each local and distant locus, we considered all possible relationships among locus genotype, transcript abundance, and protein abundance. We computed the BIC score for each of eight possible models. For each protein, we recorded the optimal local and distant locus model (that is, model that yields the lowest BIC score). In addition, we calculated the Bayesian posterior probability (assuming a uniform prior over relationships), and from these posterior probabilities estimated the expected number of proteins for each model.

**Mediation analysis to identify distant regulators and co-regulated proteins.** For proteins with distant pQTL, mediation analysis was used to identify proteins and transcripts in that region that were likely to be the causal mediator of the QTL. Mediation analysis in this context is adapted from the general approach outlined previously<sup>50</sup> to differentiate moderator from mediator variables in social psychology research<sup>51</sup>. We implemented our method as the function 'intermediate' for the open statistical language R. In brief, for a given distant pQTL, we first identified all expressed proteins and transcripts within 10 Mb of the peak SNP—these genes are candidate mediators of the distant pQTL. We then included the protein abundance of each candidate individually as an additive covariate in the pQTL mapping model and re-ran the regression at the peak distant SNP. We performed the same analysis with transcript abundance as the additive covariate. Our expectation was that many distant pQTL would be mediated by the protein and/or transcript abundance of a gene in that locus. For distant pQTL where this is true, including the abundance of the mediator protein/transcript in the pQTL mapping model should significantly decrease or abolish the distant pQTL effect—as evidenced by a decrease in LOD score. We calculate LOD scores using the 'double-lod-diff' method in *r/intermediate* to minimize the effects of missing data in the proteomics and RNA-seq data sets.

**Statistical analysis.** To assess the significance of the LOD drop for a given candidate mediator on a given distant pQTL, a null distribution of LOD scores was estimated by re-running the regression at the peak SNP and including all expressed proteins and transcripts outside of the candidate regions as additive covariates. In total, this yields mediation LOD scores for 8,050 proteins and 21,454 transcripts for each distant pQTL. Mediation LOD scores are then scaled to *z*-scores, and any candidate with a conservative *z*-score  $\leq -6$  is recorded as a potential causal mediator. Further, any protein/transcript outside of the pQTL window with a *z*-score  $\leq -6$  is recorded as a potential co-regulated partner of the target protein. We examined 1,130 distant pQTL and identified at least one candidate protein or transcript mediator for 743. In total, we found 618 unique protein/transcript mediators, of which 534 regulated a single protein, 61 regulated two proteins, and 23 regulated three or more proteins. Furthermore, 84% of the top candidate protein mediators were themselves driven by a local pQTL.

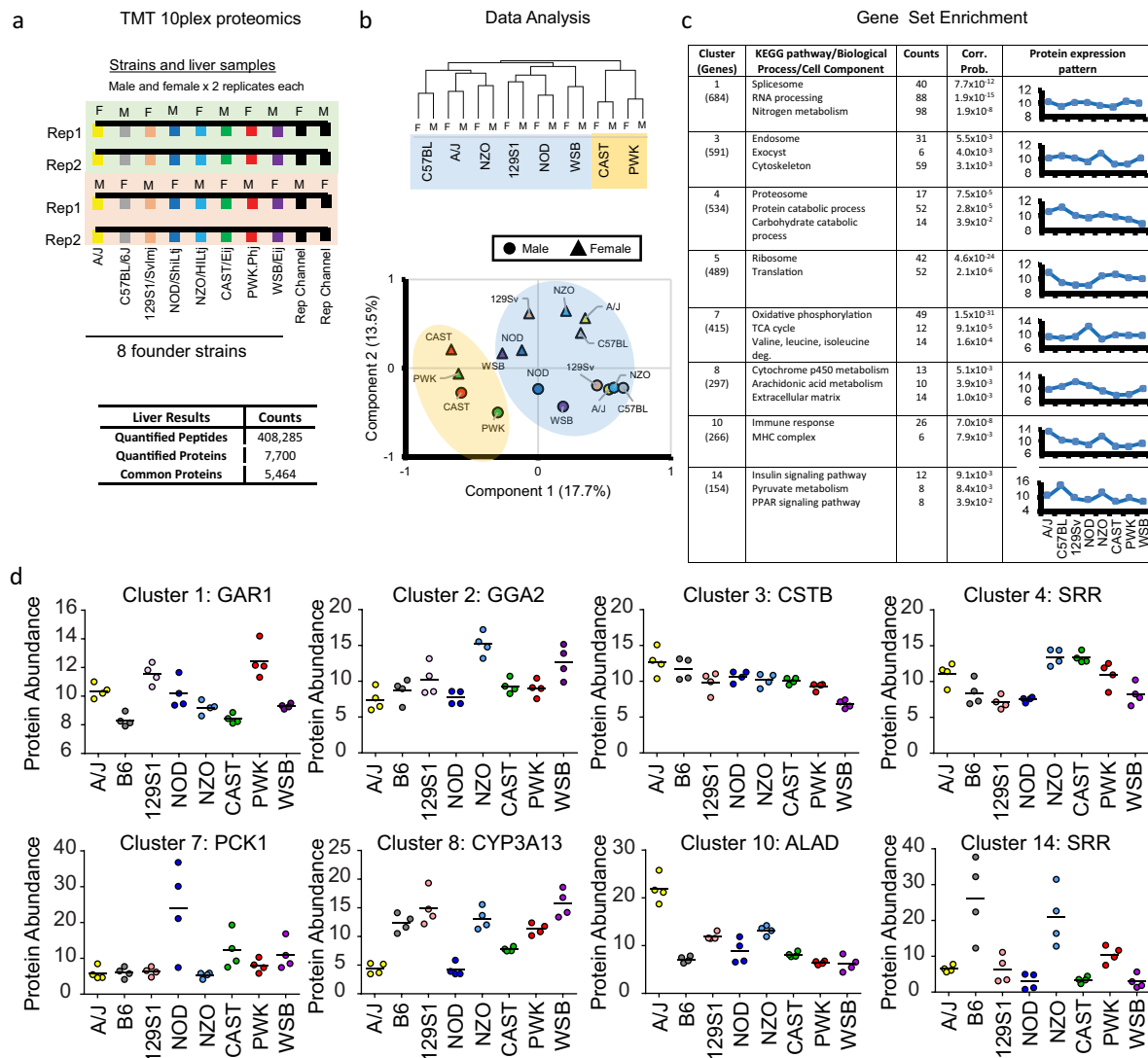
**Analysis of distant pQTL for transcriptional modes of regulation.** We observed that a small subset of local pQTL and nearly all distant pQTL lacked corresponding eQTL. For these proteins, transcript and protein abundance appeared to be largely uncoupled (buffered). For the minority of local pQTL lacking corresponding local eQTL, we expected that mutations altering protein stabilization but not affecting transcript abundance conferred this effect. The paucity of distant pQTL with corresponding eQTL is especially puzzling given our initial expectation that *trans* effects on protein abundance would likely stem

from transcription factors or chromatin modifying proteins. We detected few transcription factors and fewer transcription factor pQTL in our protein data set ( $n = 132$  expressed out of 2,243 annotated transcription factors;  $n = 21$  out of 132 transcription factors with pQTL;  $n = 9$  local transcription factor pQTL,  $n = 12$  distant transcription factor pQTL), suggesting (as others have noted<sup>52</sup>) that their regulation is more evolutionarily constrained and less tolerant of genetic variation, or alternatively, that the effects of any individual polymorphism in a transcription factor may be buffered by other transcriptional components. Results from recent large population genetics data sets<sup>53</sup> support the former explanation, and consequently distant effects from transcription factors may resist detection by genetic mapping methods and account for the lack of distant pQTL that affect both transcript and protein abundance.

**Assembly and clustering of the distant pQTL regulatory network.** We assembled the distant pQTL regulatory network by drawing directed edges to connect each *trans*-pQTL with its primary target protein. Each target protein was then connected to co-regulated proteins via directed edges. For purposes of graph assembly, each distant pQTL was represented by the protein most likely to be responsible for the effects of the QTL as indicated by mediation analysis. To identify clusters of co-regulated proteins, the directed network was converted to undirected form and subjected to MCL clustering<sup>54</sup> using an inflation parameter of 1.5. Each cluster was then evaluated for enrichment of PFAM domains<sup>55</sup>, subcellular localizations<sup>56</sup>, or GO categories<sup>57</sup> using a hypergeometric test with subsequent multiple testing correction<sup>58</sup>.  $P < 0.05$  after multiple testing correction was considered indicative of enrichment.

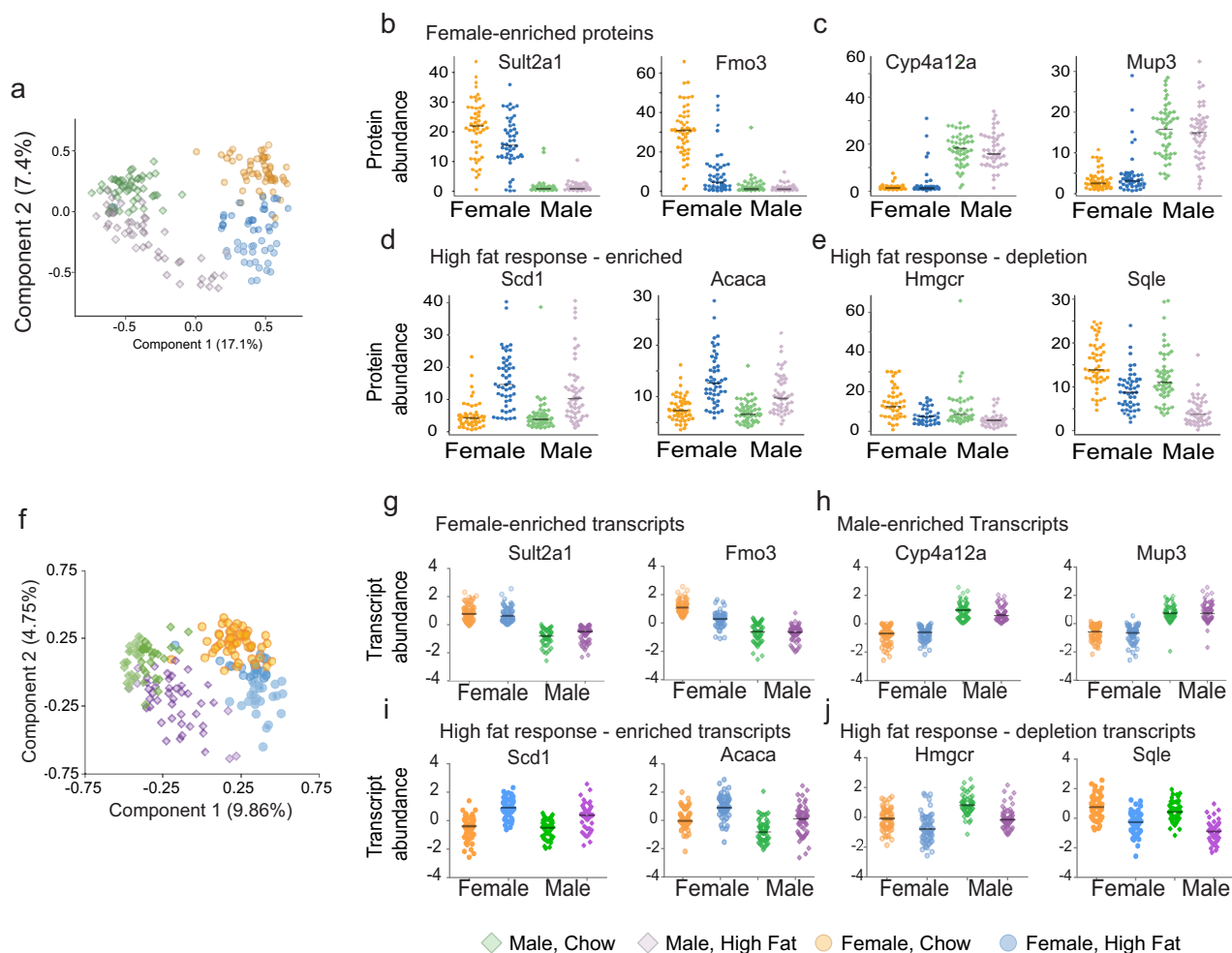
**Mapping distant pQTL and co-regulated proteins onto the BioPlex protein interaction network.** To quantify the extent to which direct physical interactions could explain distant pQTL regulation, each distant pQTL and its regulated proteins were associated with their human homologues using official gene symbols and mapped to the BioPlex network of human protein interactions<sup>31</sup>. Any protein that could not be mapped to the BioPlex network, either because a human homologue was not known or because the protein did not occur in the network, was excluded. Physical interactions connecting the pQTL and its co-regulated proteins were counted and compared against the maximum number of pairwise connections to calculate the density of physical interactions. A binomial model was used to identify sets with unusually high numbers of interactions assuming the probability of an interaction occurring between two randomly selected proteins in the BioPlex network was  $9.42 \times 10^{-4}$  (the BioPlex graph density).  $P$  values were adjusted for multiple hypothesis testing using the method of Benjamini–Hochberg<sup>58</sup> and those smaller than 0.05 after correction were taken to be significant.

41. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
42. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
43. de Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
44. Welsh, C. E. & McMillan, L. Accelerating the inbreeding of multi-parental recombinant inbred lines generated by sibling matings. *G3 (Bethesda)* **2**, 191–198 (2012).
45. Broman, K. W. *et al.* Haplotype probabilities in advanced intercross populations. *G3 (Bethesda)* **2**, 199–202 (2012).
46. Munger, S. C. *et al.* RNA-Seq alignment to individualized genomes improves transcript abundance estimates in multiparent populations. *Genetics* **198**, 59–73 (2014).
47. Cheng, R., Abney, M., Palmer, A. A. & Skol, A. D. QTLRel: an R package for genome-wide association studies in which relatedness is a concern. *BMC Genet.* **12**, 66 (2011).
48. Dudbridge, F. & Koeleman, B. P. C. Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies. *Am. J. Hum. Genet.* **75**, 424–435 (2004).
49. Storey, J. D., Taylor, J. E. & Siegmund, D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Series B* **66**, 187–205 (2004).
50. Baron, R. M. & Kenny, D. A. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51**, 1173–1182 (1986).
51. Fritz, M. S. & Mackinnon, D. P. Required sample size to detect the mediated effect. *Psychol. Sci.* **18**, 233–239 (2007).
52. Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35**, 57–64 (2003).
53. Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709 (2013).
54. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
55. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
56. Magrane, M. & UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* **2011**, bar009 (2011).
57. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
58. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing on JSTOR. *J. R. Stat. Soc. B* **57**, 289–300 (1995).



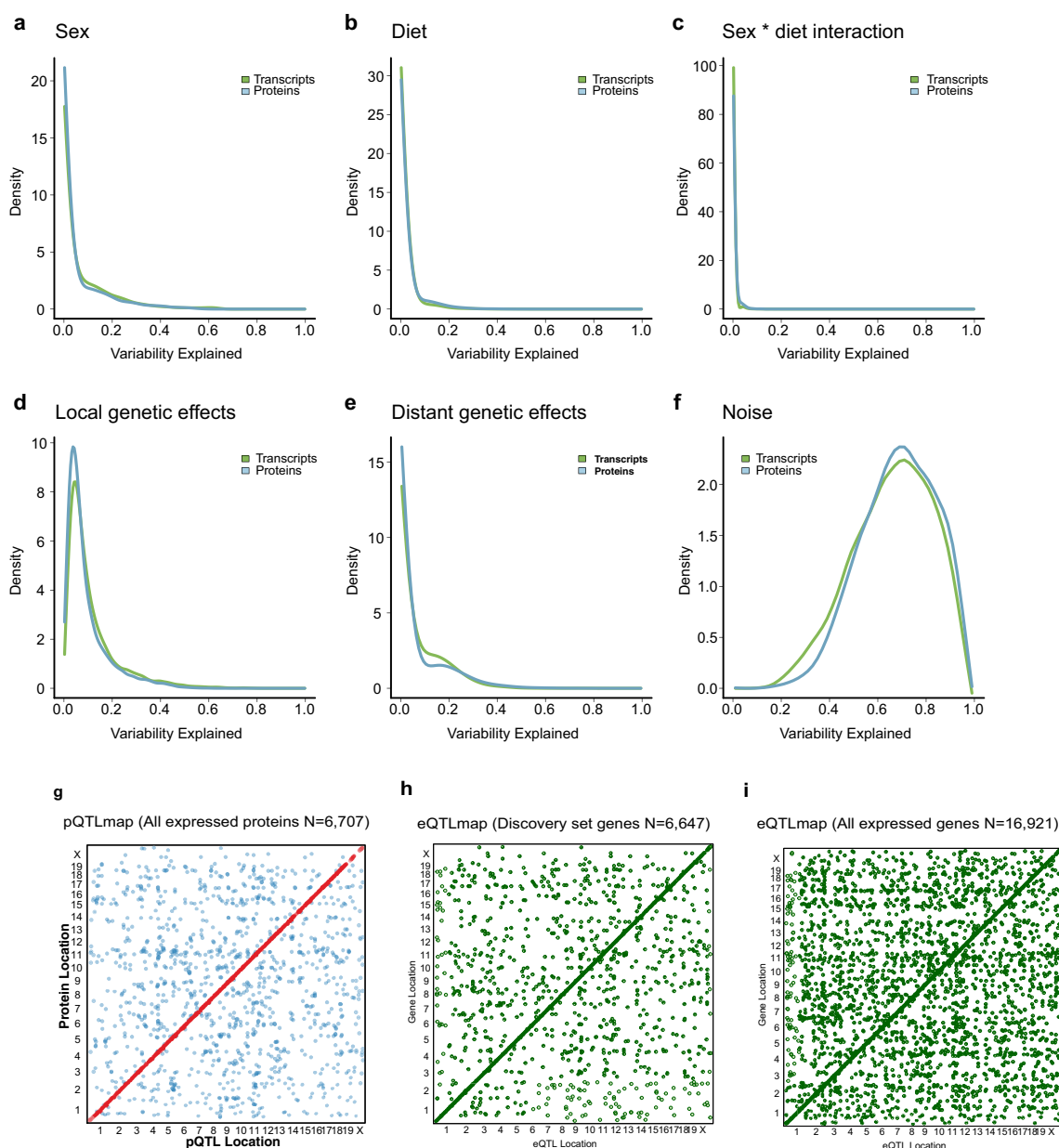
**Extended Data Figure 1 | Proteomic profiling of the eight founder strains used to create the DO mouse population.** **a**, A multiplexed TMT proteomics method was used to characterize protein expression for the eight founder strains with two biological replicates for each strain using both sexes. In total, just over 400,000 peptides were quantified corresponding to 7,699 proteins. **b**, Hierarchical clustering and principal component analysis determined that the major source of variation in protein expression is due to genetic variation among the eight strains and the sex within strains. **c**, K-means clustering and gene set enrichment

determined that each of the clusters was specifically enriched for metabolic pathways, biological process or cellular components. **d**, Proteins representing each of the displayed clusters from **c**. These proteins have specific patterns of expression as exemplified by PCK1, which was highly expressed in the NOD strain. Other examples include SCD1, which was highly expressed in C57BL/6J and NZO strains ( $n = 4$  mice for each founder, 2 male and 2 female, black bars represent median values). Protein abundance is shown as the percentage contribution of that mouse's protein levels to its respective 10-plex.



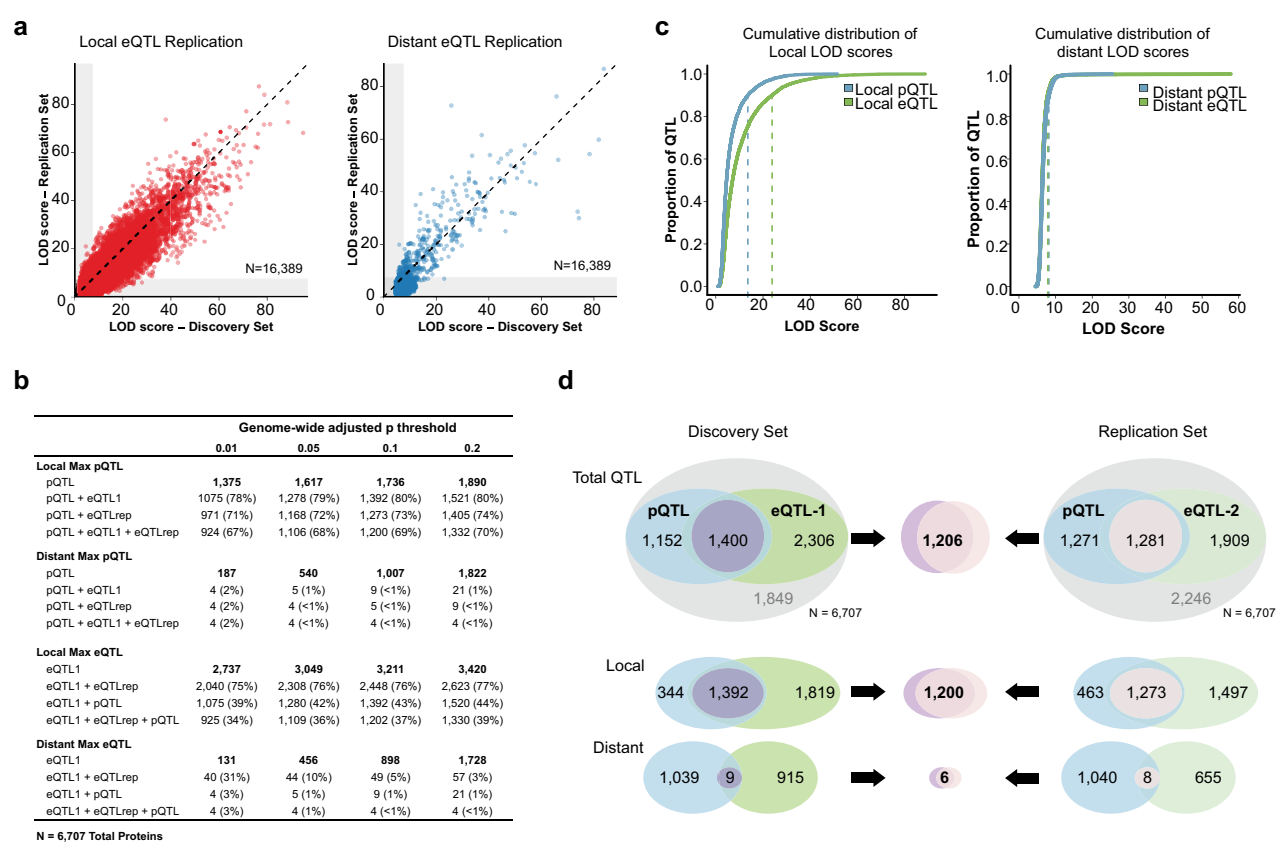
**Extended Data Figure 2 | The influence of sex and diet on protein and transcript abundance.** **a**, Principal component analysis aligns well with sex and diet as major experimental contributors of variation in protein abundance. **b**, Female-specific protein abundance profiles for SULT2A1 and FMO3. **c**, Male-specific protein abundance profiles for CYP4A12A and MUP3. **d**, **e**, Diet also resulted in the regulation of many proteins,

which are represented by proteins such as SCD1 and ACACA that increased in abundance and proteins such as HMGCR and SQLE that decreased in abundance. **f**, Principal component analysis aligns well with sex and diet as major experimental contributors of variation in transcript abundance. **g**–**j**, Transcript scatter plots for the proteins in **b**–**e**. Transcript abundance data were transformed to rank normal scores for plotting.



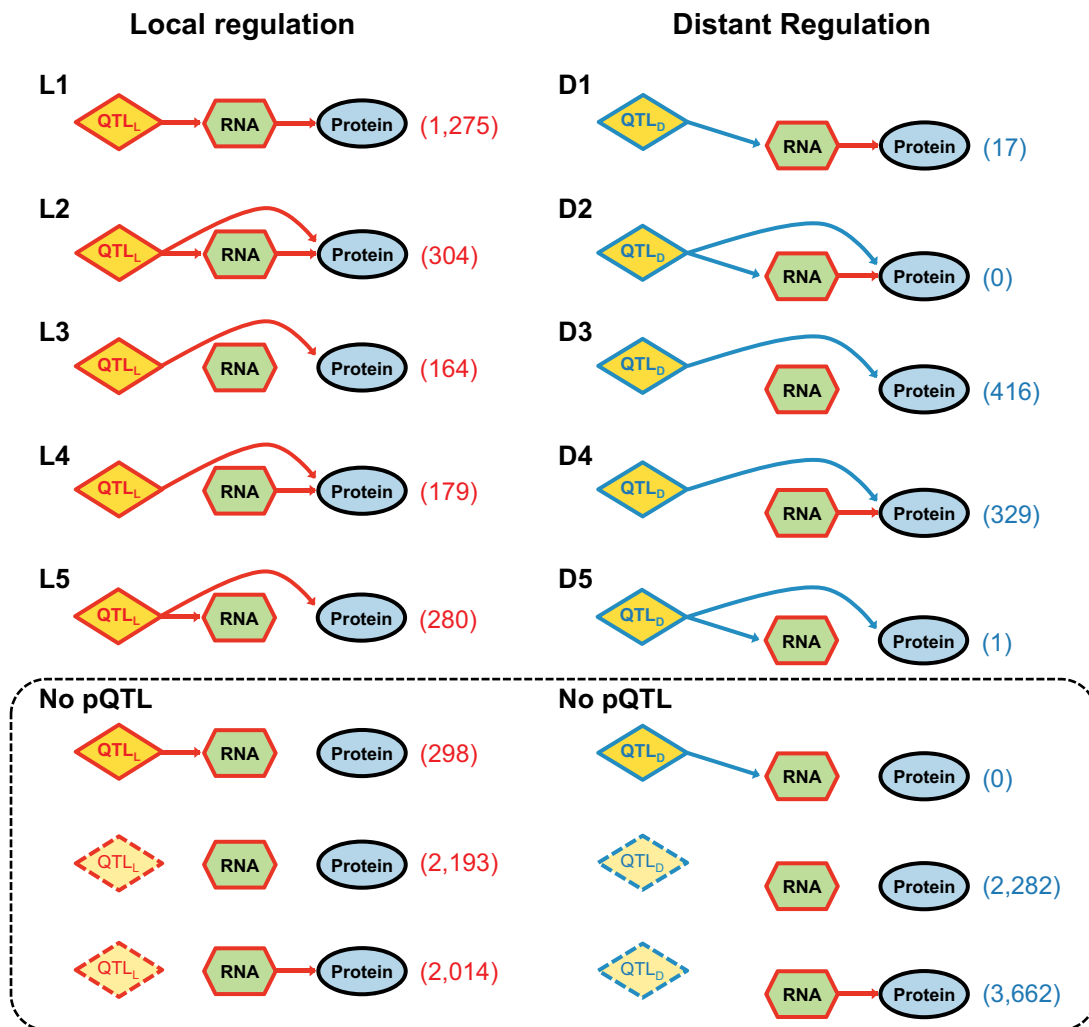
**Extended Data Figure 3 | Genetic effects drive much of the observed expression variance in the RNA-seq and proteomics data.** Liver transcript and protein abundance are highly variable in the DO population. Among the discovery set ( $n = 6,707$  proteins, 6,647 genes), much of this variance can be attributed to one or more experimental variables and/or genetic effects. **a–c**, The experimental covariates sex and diet influence many transcripts and proteins in an additive manner, however, the interaction of sex and diet does not seem to affect many genes. The effects from sex and diet are not biased towards one molecular species—that is, similar numbers of transcripts and proteins are similarly affected by these experimental variables. Genetic variation underlies many of the most variable transcripts and proteins. **d, e**, Local genetic variation in particular is a strong driver of expression variation for many

genes, while distant genetic effects are observed but more subtle. Among the discovery set, we observe more and larger genetic effects (both local and distant) on transcript abundance than protein abundance. **f**, For most transcripts and proteins detected in this study, expression variation is minimal, cannot be attributed to a known experimental or genetic variable, and is plotted as noise. **g**, pQTL map for all 6,707 proteins tested from genetic linkage analysis. **h, i**, QTL mapping identified the genetic loci that underlie variability in transcript abundance (eQTL). For the discovery set of transcripts with detected proteins and the larger set of all expressed genes, the location of the eQTL is plotted on the x axis and the location of the controlled gene is plotted on the y axis. Most genetic effects are local and map to the same location as the gene, as evidenced by the prominent diagonal line in both maps.



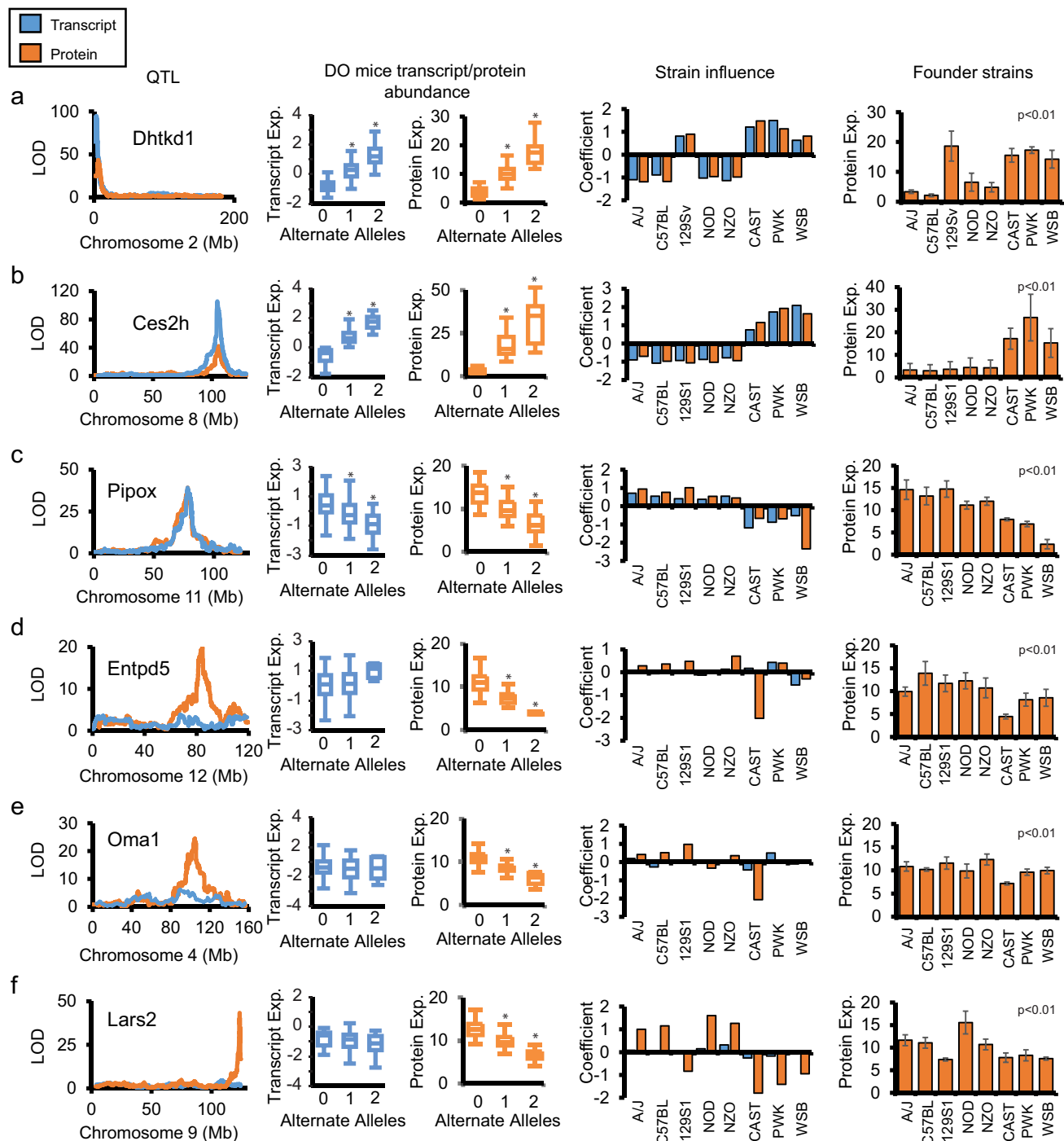
**Extended Data Figure 4 | Replication rates for eQTL are highly correlated with effect size, and local eQTL replicate at higher rates than distant eQTL.** **a**, To assess replication of eQTL, an independent set of 192 DO liver RNA-seq samples was analysed ('replication set') and compared to the discovery set. A total of 16,839 genes were expressed in half or more samples in both data sets. For each gene, the most significant proximal locus (within  $\pm 10$  Mb of gene) and distant locus (located on a different chromosome from the gene) were identified from the discovery set—LOD scores at these loci are plotted on the x axis (local in red; distant in blue). Next, the most significant loci within a 10-Mb window flanking the local and distant loci from the discovery set were identified in the replication set and plotted on the y axis. LOD scores are highly correlated at these peak loci (local Pearson  $r = 0.91$ ; distant  $r = 0.84$ ). **b**, For the core set of 6,707 proteins (6,647 gene ids), pQTL and eQTL overlap were compared at multiple genome-wide  $P$  value thresholds from 0.01 to 0.2. Again, one maximum proximal locus and one maximum distant locus were identified for each gene/protein, and recorded if it met the  $P$  value cut off. Local pQTL exhibit high overlap with both the discovery eQTL set and replication eQTL set, regardless of  $P$  value threshold (67–80%). Distant pQTL exhibit slightly higher overlap with eQTL at the most stringent  $P$  value cut off, however, overlap is consistently low for distant pQTL (<1–2%). Local eQTL overlap well with the replication eQTL set

regardless of  $P$  value threshold (75–77%). Distant eQTL replicate poorly overall (3–31%), but overlap rate is highest (31%) at the most stringent  $P$  value threshold, suggesting that larger sample sizes will be required to fully and accurately characterize distant effects on gene expression. **c**, The maximum proximal locus and distant locus were identified for each of the 6,707 proteins and transcripts, and the cumulative distribution of their LOD scores is plotted (blue = proteins, green = transcripts). LOD score is plotted on the x axis, and the proportion of total QTL is plotted on the y axis. Local eQTL as a group exhibit higher LOD scores (consistent with higher effect sizes) than local pQTL (ninetieth percentile LOD = 23.9 for local eQTL, 13.6 for pQTL), while distant eQTL and pQTL are of similar scale (ninetieth percentile LOD = 7.9 for distant eQTL, 8.2 for distant pQTL). **d**, Comparison of pQTL from the discovery set to eQTL from the discovery set (left set of Venn diagrams) and eQTL from the replication set (right). As expected given that they derive from the same samples, local pQTL and eQTL overlap is observed to be higher in the discovery set (1,392 out of 1,736 = 80%), however, local pQTL still overlap well with eQTL from the replication set (1,273 out of 1,736 = 73%). Distant pQTL overlap poorly with both eQTL sets (9 out of 1,048 in discovery set); 8 out of 1,048 in replication set), however, 6 of 9 distant pQTL that do overlap with eQTL in the discovery set are also identified as overlapping in the replication set.



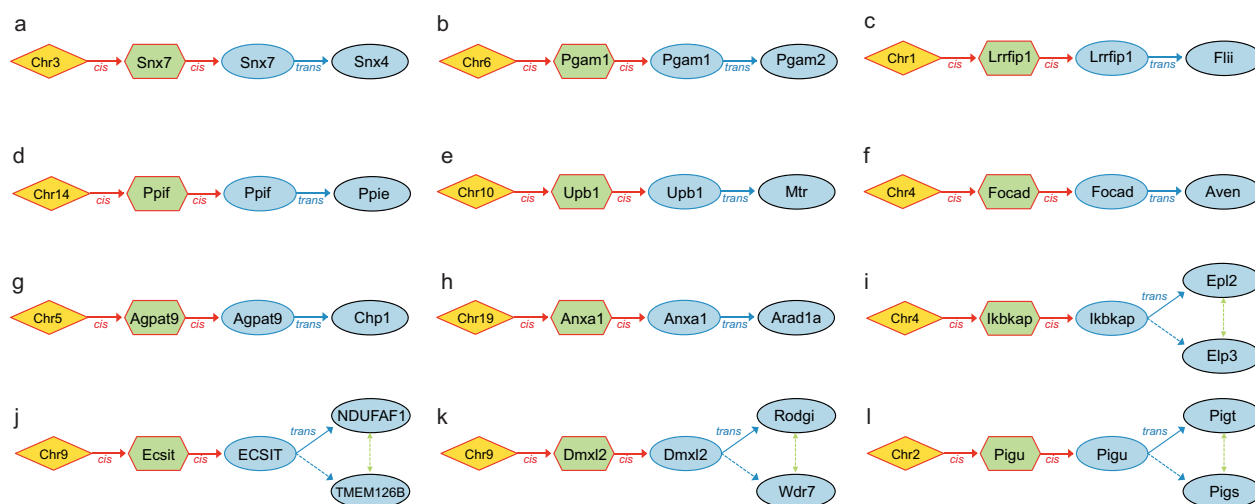
**Extended Data Figure 5 | BIC model selection reveals transcriptional mechanisms driving most local pQTL and post-transcriptional mechanisms underlying most distant pQTL.** We identified the local and distant QTL with the maximum LOD score (regardless of significance) for each of the 6,707 proteins, and used BIC to assess eight models linking QTL genotype to transcript and protein abundance. Most proteins are not affected by the local or distant QTL, and fall in one of the three groups below outlined by the dotted line. Among the five models where a QTL

effect on protein abundance is detected, two are transcriptional in nature (L1, L2; D1, D2); the QTL effect on protein abundance is conferred at least partially through the transcript. The remaining three genetic models are post-transcriptional (L3–5; D3–5); the QTL effect on protein abundance is not mediated through the transcript. The transcriptional L1 and L2 models are identified as the best models for most local pQTL, while the post-transcriptional D3 and D4 models are optimal for most distant pQTL.



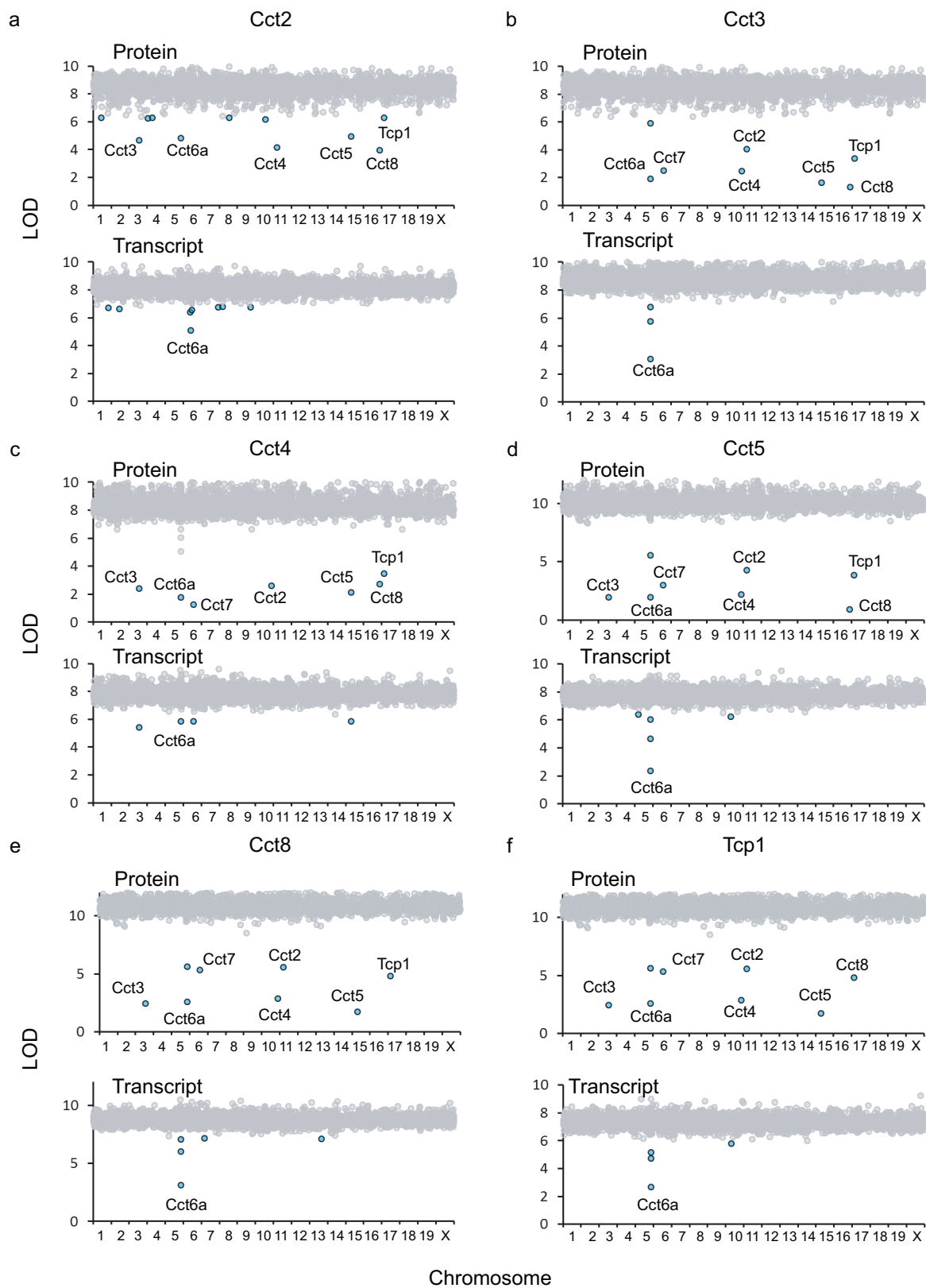
**Extended Data Figure 6 | Examples of local pQTL that are due to an underlying eQTL and those that are due to post-transcriptional mechanisms.** **a**, The protein DHTKD1 contained a local acting eQTL and pQTL, which was associated with increased transcript and protein abundance derived from 129S1/SvImJ, CAST/EiJ, PWK/PhJ and WSB/EiJ strains. Mice were divided into three groups depending on whether or not their genomes contained 0, 1 or 2 of the alleles found to be associated with the pQTL. These increases in protein abundance were further validated using the proteomic analysis of the founder strains. **b**, **c**, Similarly, *Ces2h* and *Pipox* had both a local acting eQTL and pQTL that could be associated with specific strains (CAST/EiJ, PWK/PhJ and WSB/EiJ). These protein abundance measurements were further validated using the founder strains

data set. **d**, **e**, Alternatively, 10% of the genes had local pQTL but lacked local eQTLs, which is evident in proteins such as ENTPD5 and OMA1. The founder allele expression patterns inferred at the pQTL were validated by protein abundance measurements in the founder strains, which could be explained CAST/EiJ specific missense mutations in both genes. **f**, Likewise, *Lars2* also contained a pQTL that had no observable eQTL that showed a decrease in protein abundance in the 129S1/SvImJ, CAST/EiJ, PWK/PhJ and WSB/EiJ strains. Genome sequencing determined that these strains share four missense mutations ( $*P < 0.01$  using a Student's *t*-test; for founder strains,  $n = 4$  mice for each founder, 2 male and 2 female, error bars represent s.d.).



**Extended Data Figure 7 | The causal relationship between genetic variation and protein expression was determined for over 700 proteins as inferred by mediation analysis. a–d,** Many of the causal relationships between proteins have been previously documented such as the associations between SNX7–SNX4, PGAM1–PGAM2, LRRFIP1–FLII and PPIF–PPIE. **e–h,** In addition, many of the protein associations had

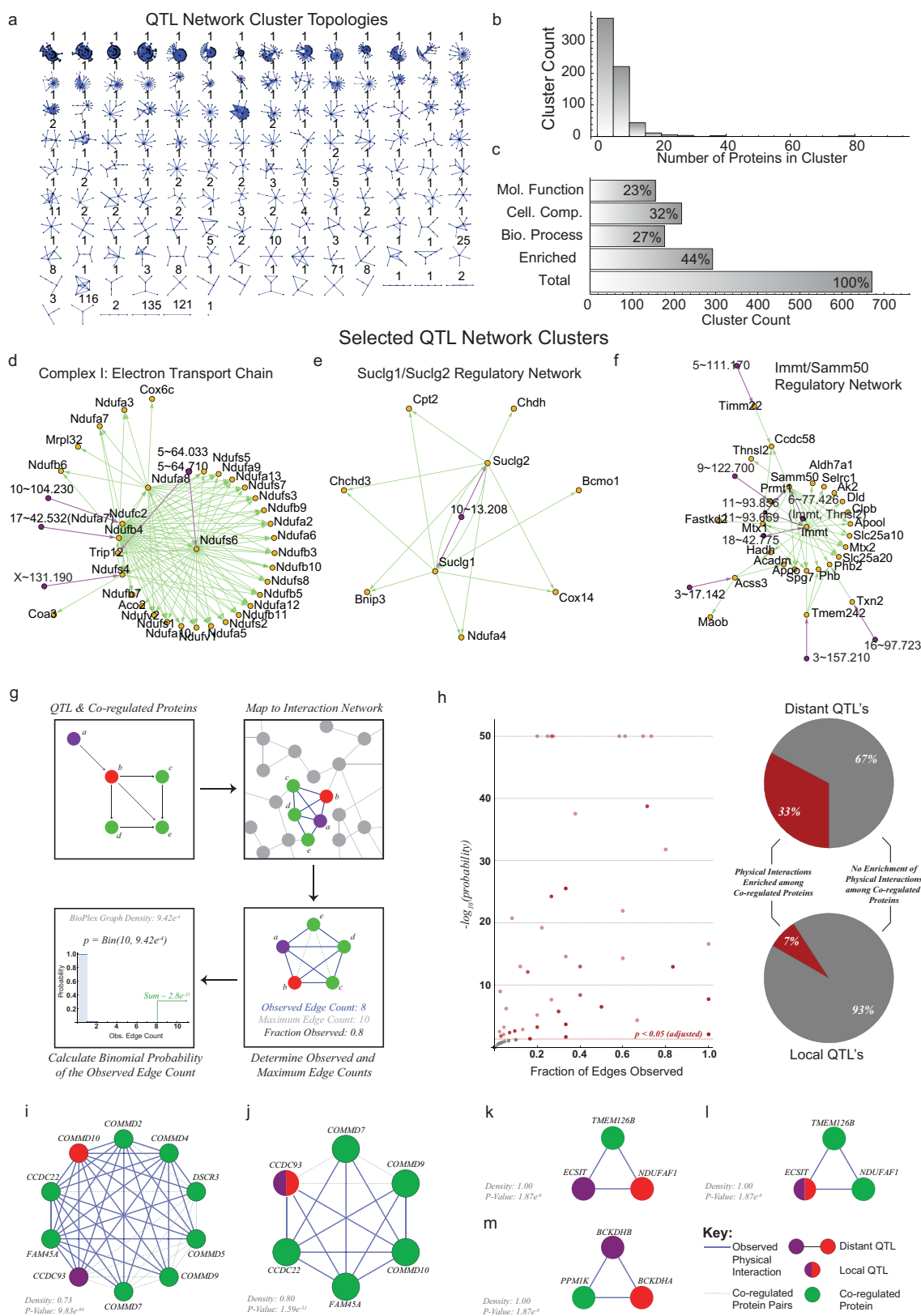
not be previously documented such as UPB1–MTR, FOCAD–AVEN, AGPAT9–CHP1 and ANXA1–ARAD1A. **i–l,** Protein associations were also identified for multimeric complexes such as ECSIT–NDUFAF1–TMEM126B, DMXL2–ROGDI–WDR7, PIGU–PIGT–PIGS and IKBKAP–ELP2–ELP3.



## Chromosome

**Extended Data Figure 8 | Mediation analysis for CCT complex members details the effects of a QTL in *Cct6a* on protein abundance through post-transcriptional protein buffering. a–f, Mediation analysis for each of the Cct complex identifies *Cct6a* as the causal intermediate. A local QTL for *Cct6a* affects transcript and protein abundance, and CCT6A**

abundance sets the abundance of other CCT proteins regardless of variation in their transcripts. For each of the complex members tested, all other complex members are confirmed to be co-regulated providing additional supporting evidence for stoichiometric buffering.

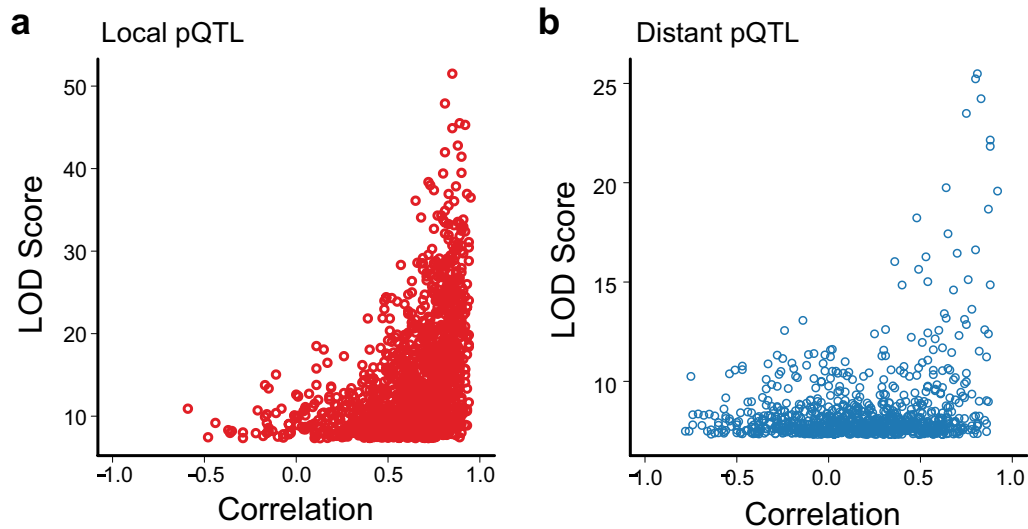


Extended Data Figure 9 | See next page for caption.

### Extended Data Figure 9 | Distant pQTL and co-regulated proteins frequently correspond to complexes of physically interacting proteins.

**a**, Distant pQTL and co-regulated proteins assemble to form a regulatory network, which is defined by protein clusters with distinct topologies. A total of 3,938 proteins/QTL are linked by 5,794 associations. Distant pQTL are depicted as purple arrows pointing from the inferred causal protein to its regulated pair. Co-regulated proteins are connected with green arrows emanating from the primary target protein. **b**, MCL clustering decomposes the distant pQTL network into 671 clusters. Cluster size varies considerably, although most clusters contain fewer than 20 proteins. **c**, Clusters extracted from the distant pQTL network frequently associate proteins with shared biological functions. More than half of clusters are enriched for at least one GO category, as depicted in the bar chart above. **d–f**, Three selected clusters of distant pQTL and co-regulated proteins. **g**, To understand the relationship between the distant pQTL associations and protein interactions, each distant pQTL and its co-regulated proteins were mapped to their human homologues in the BioPlex network of human protein interactions. To assess the tendency

for these co-regulated proteins to cluster together, the median graph distance separating all pairs of co-regulated proteins was determined. The distribution of median distances observed for equal numbers of randomly selected proteins was also determined and used to assign a Z-score to each distant pQTL and its co-regulated proteins. **h**, Histogram depicting the Z-score distribution for distant pQTL and co-regulated proteins. Z-scores below  $-2.5$  (highlighted in red) indicated that co-regulated proteins were unusually close within the BioPlex network. **i–l**, Selected distant pQTL and co-regulated proteins, mapped onto the BioPlex network of protein interactions. All shortest paths connecting distant pQTL and their regulated proteins have been extracted from the BioPlex network and displayed. Proteins inferred to be responsible for each QTL are purple, while primary regulated proteins are red and secondary co-regulated proteins are green. Grey circles represent neighbouring proteins in the BioPlex network that were not found to be co-regulated. Grey edges indicate BioPlex interactions, while Blue edges denote co-regulation uncovered from trans-QTL analysis.



**Extended Data Figure 10 | Comparison of protein abundance in the DO and founder strains reveals a positive correlation between pQTL significance and predictive power. a, b,** For all detected liver pQTL in the DO population, founder strain allelic contributions were derived from the mapping model and compared to protein abundance measured directly from the eight founder strains. Pearson correlations are plotted against

the LOD score of the pQTL for both local and distant pQTL. Predictive power tracks well with pQTL significance. Local pQTL tend to be more significant and yield higher predictive power than distant pQTL, however highly significant distant pQTL (>10 LOD) have comparable predictive power to local pQTL of similar significance.

# Crystal structure of the epithelial calcium channel TRPV6

Kei Saotome<sup>1\*</sup>, Appu K. Singh<sup>1\*</sup>, Maria V. Yelshanskaya<sup>1</sup> & Alexander I. Sobolevsky<sup>1</sup>

**Precise regulation of calcium homeostasis is essential for many physiological functions. The  $\text{Ca}^{2+}$ -selective transient receptor potential (TRP) channels TRPV5 and TRPV6 play vital roles in calcium homeostasis as  $\text{Ca}^{2+}$  uptake channels in epithelial tissues. Detailed structural bases for their assembly and  $\text{Ca}^{2+}$  permeation remain obscure. Here we report the crystal structure of rat TRPV6 at 3.25 Å resolution. The overall architecture of TRPV6 reveals shared and unique features compared with other TRP channels. Intracellular domains engage in extensive interactions to form an intracellular ‘skirt’ involved in allosteric modulation. In the  $\text{K}^+$  channel-like transmembrane domain,  $\text{Ca}^{2+}$  selectivity is determined by direct coordination of  $\text{Ca}^{2+}$  by a ring of aspartate side chains in the selectivity filter. On the basis of crystallographically identified cation-binding sites at the pore axis and extracellular vestibule, we propose a  $\text{Ca}^{2+}$  permeation mechanism. Our results provide a structural foundation for understanding the regulation of epithelial  $\text{Ca}^{2+}$  uptake and its role in pathophysiology.**

The TRP channels are a superfamily of cation-permeable ion channels that are widely known for their role as transducers of sensory modalities<sup>1</sup>. TRPV5 and TRPV6 are TRP channels that are uniquely selective for  $\text{Ca}^{2+}$  (permeability ratio  $P_{\text{Ca}}/P_{\text{Na}} > 100$ ) (ref. 2). They have not been reported to be responsive to temperature, tastants or odours, but the mechanosensitive properties of TRPV6 appear to be important for the formation of microvilli<sup>3</sup>. TRPV5 and TRPV6 belong to the vanilloid subfamily of TRP channels, share ~75% sequence identity and are involved in the transport of calcium through epithelial cell membranes<sup>4</sup>. Knockout of TRPV6 in mice leads to various phenotypes linked to impaired  $\text{Ca}^{2+}$  homeostasis, including defective intestinal  $\text{Ca}^{2+}$  absorption, lower body weight, impaired fertility and dermatitis<sup>5–8</sup>. Altered TRPV6 expression has also been shown in various transgenic mouse models of human diseases<sup>8</sup>, including Crohn’s and kidney stone diseases. In addition, TRPV6 is implicated in the development and progression of numerous forms of cancer, and its overexpression pattern correlates with the aggressiveness of the disease<sup>9,10</sup>. Accordingly, TRPV6 has emerged as a target for diagnosing and treating various carcinomas<sup>11,12</sup>.

Structurally, TRPV5 and TRPV6 represent homo- or heteromeric assemblies of four subunits<sup>13</sup>, each containing a central  $\text{K}^+$ -channel-like transmembrane domain that is flanked by intracellular amino (N)- and carboxy (C)-terminal domains. The overall architecture and potential gating mechanisms of TRP channels have recently been illuminated by cryo-electron microscopy structures of TRPV1 (refs 14, 15), TRPV2 (ref. 16) and TRPA1 (ref. 17). However, the absence of structural bases for the unique physiological properties of TRPV5 and TRPV6 motivated us to study these epithelial  $\text{Ca}^{2+}$  channels.

## Structure determination

We screened various orthologues of TRPV5 and TRPV6 and discovered rat TRPV6 as a promising candidate for our structural studies. We modified the 727-residue wild-type rat TRPV6 polypeptide to create the crystallization construct TRPV6<sub>cryst</sub> (see Methods). Experiments with the fluorescent  $\text{Ca}^{2+}$  indicator Fura-2 AM show that cells expressing TRPV6<sub>cryst</sub> exhibit  $\text{Ca}^{2+}$  permeability similar to wild type (Extended Data Fig. 1).

The best crystals of TRPV6<sub>cryst</sub> diffracted to 3.25 Å resolution. We solved the TRPV6<sub>cryst</sub> structure by molecular replacement, and the electron density map (Extended Data Fig. 2) was readily interpretable for most of the polypeptide (see Methods). Sequence registry was aided by anomalous difference Fourier maps highlighting natural sulfur atoms of cysteines and methionines, and selenium atoms in protein with selenomethionines substituted for methionines (Extended Data Fig. 3). The resulting model of TRPV6<sub>cryst</sub> was refined to good crystallographic statistics and stereochemistry (Extended Data Table 1).

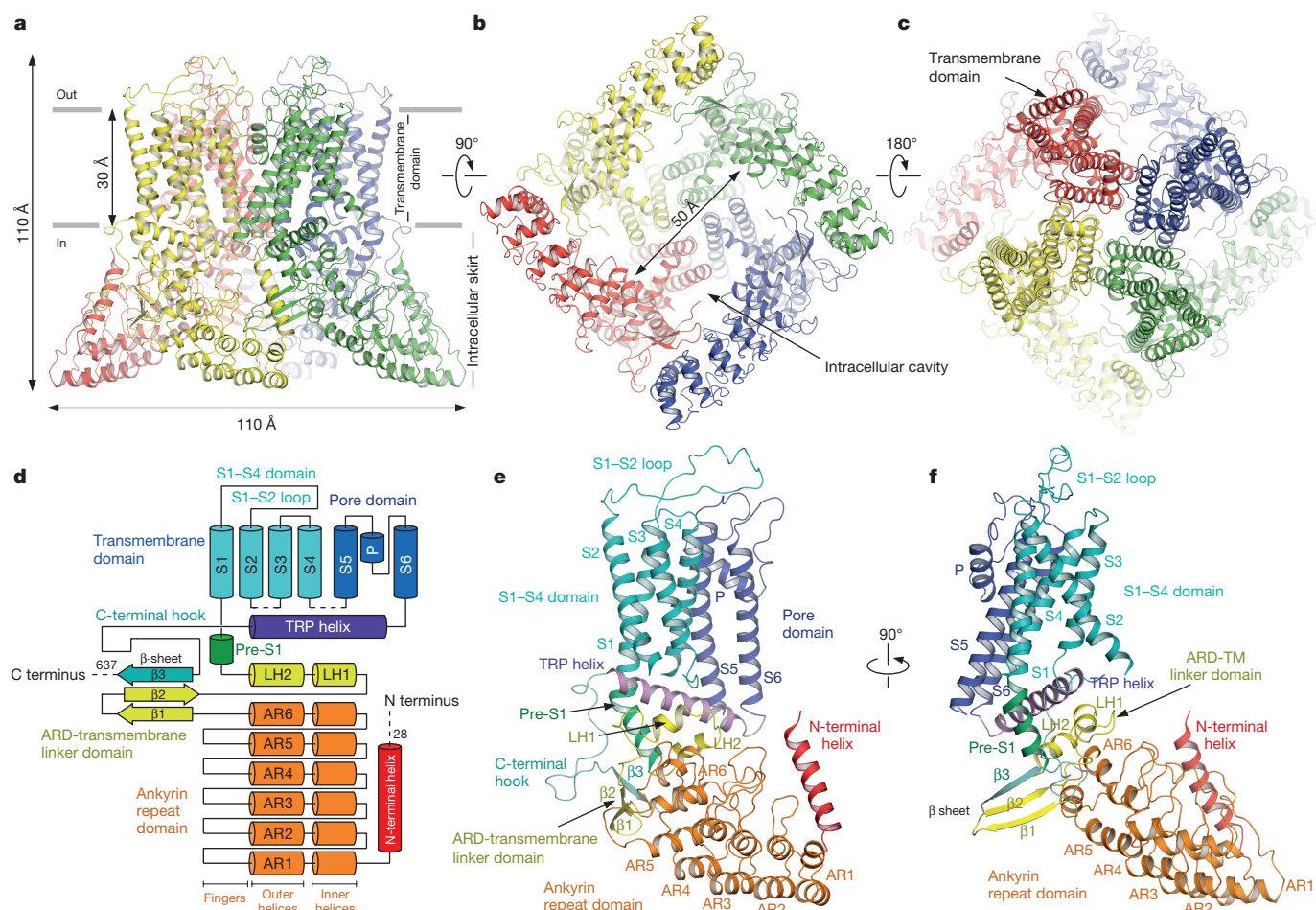
## Architecture and domain organization

The four-fold symmetrical structure of TRPV6<sub>cryst</sub> (Fig. 1) contains two main components: a transmembrane domain with a central ion channel pore, and a ~70 Å-tall and ~110 Å-wide intracellular skirt where four subunits constitute walls enclosing a ~50 Å × 50 Å-wide cavity underneath the ion channel. Like TRPV1 (ref. 14) (Extended Data Fig. 4) and TRPV2 (ref. 16), the intracellular domains of a single TRPV6<sub>cryst</sub> subunit contain an ankyrin repeat domain with six ankyrin repeats, followed by a linker domain that includes a  $\beta$ -hairpin (composed of  $\beta$ -strands  $\beta 1$  and  $\beta 2$ ) and a helix–turn–helix motif resembling a seventh ankyrin repeat, and the pre-S1 helix, which connects the linker domain to the transmembrane domain (Fig. 1d–f and Extended Data Fig. 4). Also similar to TRPV1/2, a six-residue stretch at the C terminus constitutes a  $\beta$ -strand ( $\beta 3$ ) that tethers to the  $\beta$ -hairpin in the linker domain to create a three-stranded  $\beta$ -sheet. In addition to the conserved domains, TRPV6<sub>cryst</sub> also includes an N-terminal helix and C-terminal hook, which pack against each other to form an intersubunit interface along the corners of the intracellular skirt.

Similar to other TRP channels<sup>14,16,17</sup>, the transmembrane domain of TRPV6<sub>cryst</sub> crudely resembles voltage-gated  $\text{K}^+$  (ref. 18) or  $\text{Na}^+$  (ref. 19) channels and includes six transmembrane helices (S1–S6) and a pore loop (P-loop) between S5 and S6. The first four transmembrane helices form a bundle to constitute the S1–S4 domain. The packing of aromatic side chains in S1–S4 rigidifies the helical bundle conformation (Extended Data Fig. 4c), suggesting that this domain remains relatively static during gating. The linker between the S1–S4 domain and pore

<sup>1</sup>Department of Biochemistry and Molecular Biophysics, Columbia University, 650 West 168th Street, New York, New York 10032, USA.

\*These authors contributed equally to this work.



**Figure 1 | Architecture and domain organization of TRPV6<sub>cryst</sub>.** **a–c**, Side (**a**), bottom (**b**) and top (**c**) views of the TRPV6<sub>cryst</sub> tetramer, with each subunit shown in a different colour. **d**, Domain organization diagram of the TRPV6 subunit. **e**, **f**, Two views of the TRPV6<sub>cryst</sub> subunit, with domains coloured as in **d**.

domain is unstructured, which is a marked contrast from other TRP channels, in which it assumes a helical conformation and mediates interdomain interactions<sup>14,16,17</sup>. Following S6 is the amphipathic TRP helix, which runs parallel to the membrane and interacts with intracellular soluble domains in a manner analogous to TRPV1, TRPV2 and TRPA1 (refs 14, 16, 17).

Although the overall domain organization of TRPV6<sub>cryst</sub> resembles TRPV1/2 (refs 14, 16) and, to a lesser degree, TRPA1 (ref. 17), electron density for the linker between S6 and TRP helix (Extended Data Fig. 2f) and disulfide crosslink experiments (Extended Data Fig. 5a–c) imply a unique non-swapped transmembrane domain arrangement in which the S1–S4 domain and pore domain of the same protomer are packed against each other. While this unique domain arrangement could have profound implications, we present this aspect of the TRPV6<sub>cryst</sub> model cautiously because of the absence of interpretable density for the S4–S5 linker.

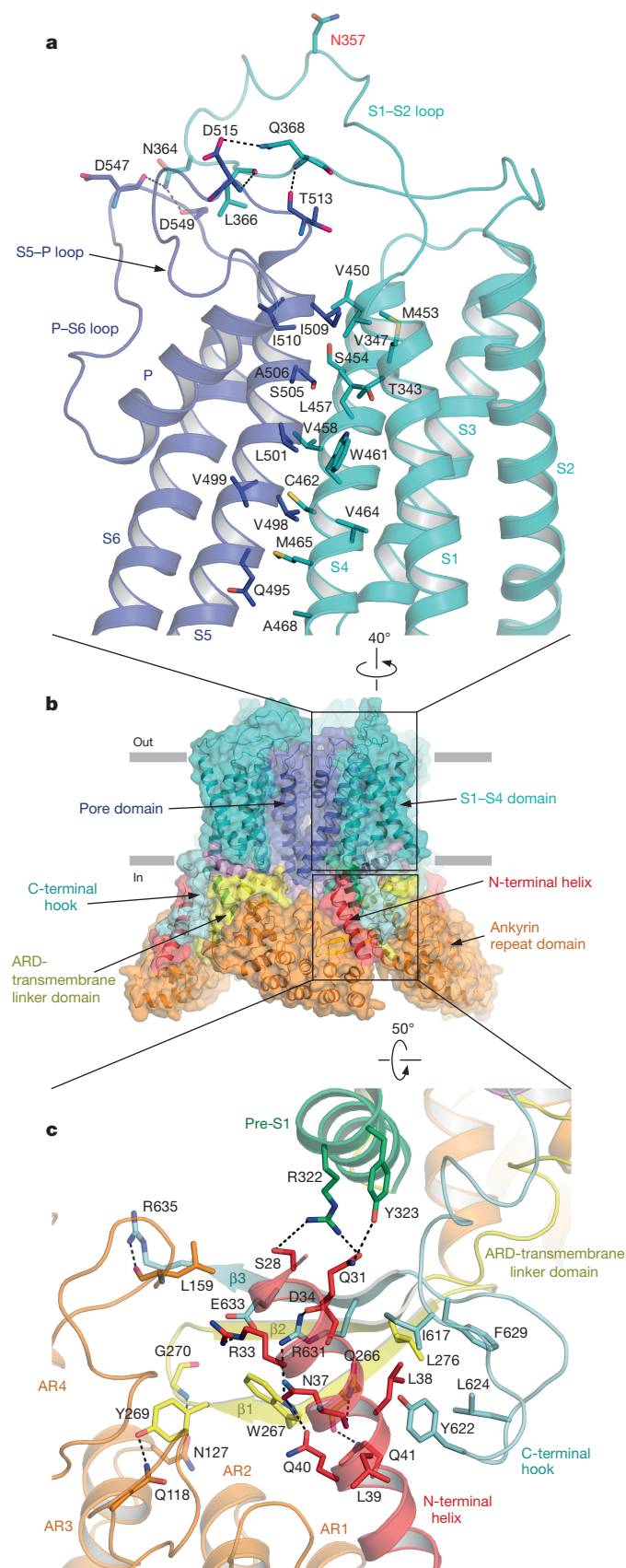
### Assembly and subunit interfaces

Assembly of TRPV6<sub>cryst</sub> is mediated by multiple interdomain interfaces (Fig. 2). Close packing of S5 against S4 and S1 of the adjacent S1–S4 domain immobilizes the pore module with respect to the S1–S4 domain (Fig. 2a), a trait that is reminiscent of the Slo2.2 K<sup>+</sup> channel<sup>20</sup> and distinct from voltage-gated channels<sup>18</sup>. Further, the S1–S2 extracellular loop contacts the S5–P and P–S6 loops (Fig. 2a). This interaction hints at a structural basis for the regulation of TRPV5 and TRPV6 function by the  $\beta$ -glucuronidase klotho, which modulates channel activity by modifying the conserved N-linked glycosylation site<sup>21</sup> located in the middle of this loop (N357 in TRPV6<sub>cryst</sub>).

The intracellular domains of TRPV5 and TRPV6 have been implicated in tetrameric assembly<sup>22</sup>, trafficking<sup>23</sup> and regulation of channel activity by the Ca<sup>2+</sup> sensor calmodulin<sup>24–26</sup>. The structure of TRPV6<sub>cryst</sub> reveals that numerous non-contiguous intracellular domains engage in extensive inter- and intrasubunit interactions (Fig. 2c). At the centre of these interactions is the N-terminal helix, which is positioned as a pillar along the corners of the intracellular skirt. Putative hydrogen bonds and salt bridges involving D34 stabilize the interaction between the N-terminal helix and three-stranded  $\beta$ -sheet. Notably, mutation of the equivalent D34 to alanine abolished Ca<sup>2+</sup> uptake function in TRPV5 (ref. 23), suggesting this interaction's functional importance. The N-terminal helix also forms hydrophobic and hydrogen bonding interactions with the C-terminal hook and pre-S1 helix from an adjacent subunit. Since it is a hub for domain interactions, endogenous or exogenous factors could allosterically modulate channel activity by targeting the N-terminal helix. Interestingly, we observed a robust cylindrical density at the intersubunit interface formed by the N-terminal helix, ankyrin repeat domain and three-stranded  $\beta$ -sheet (Extended Data Fig. 5d–h). We have tentatively attributed this density to desthiobiotin (DTB), which was included as an eluent in the TRPV6<sub>cryst</sub> affinity purification procedure (see Methods).

### Ion-conducting pore

The extracellular portion of the TRPV6<sub>cryst</sub> ion-conducting pore is formed by extracellular loops connecting the P-loop helix to S5 and S6, while the rest of the ion conduction pathway is formed entirely by the S6 helices (Fig. 3). Such pore architecture is conserved over the entire family of tetrameric ion channels (Extended Data Fig. 6).



The region connecting S5 and S6 contains eight acidic residues per protomer, four of which face the ion conduction pathway to produce a highly electronegative 'mouth' to the pore (Fig. 3a–c). Below this extracellular vestibule is a four-residue selectivity filter (<sup>538</sup>TIID<sup>541</sup>)

**Figure 2 | Domain interfaces.** **a**, Transmembrane helices S4 and S5 and extracellular loops S1–S2, S5–P and P–S6 contribute to interfaces between the S1–S4 domain and pore domain. **b**, Side view of the TRPV6<sub>cryst</sub> tetramer with domains coloured differently. Boxes indicate domain interfaces expanded in **a** and **c**. **c**, Interfaces between soluble domains. Residues at domain interfaces in **a** and **c** are shown in stick representation, with potential hydrogen bonds and electrostatic interactions shown as dashed lines. The predicted N-linked glycosylation site conserved in TRPV6 channels, N357 in TRPV6<sub>cryst</sub> is labelled red in **a**.

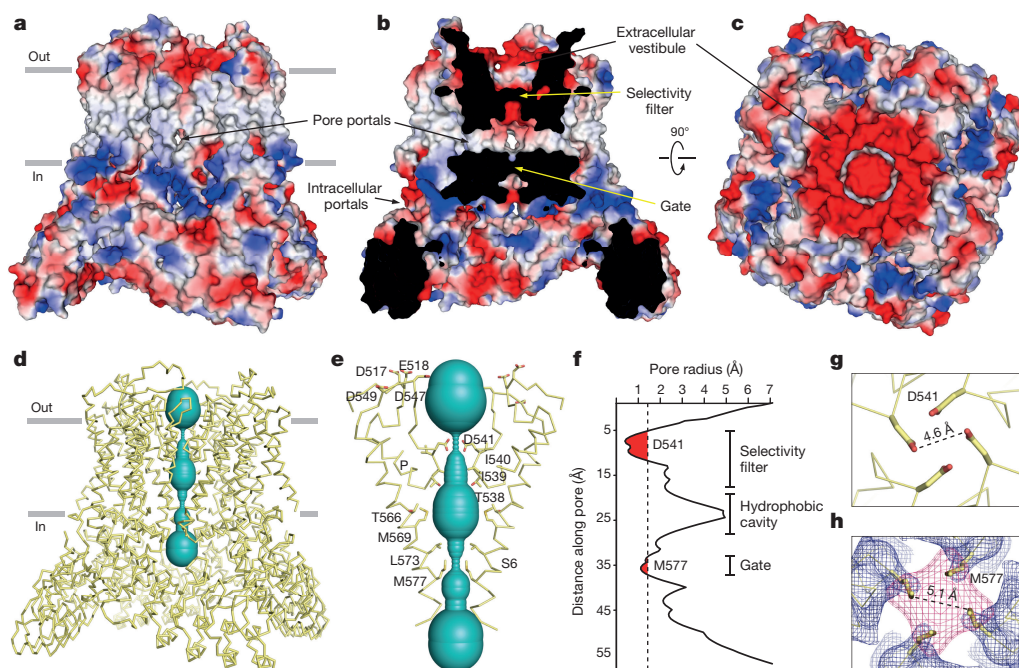
(Fig. 3d–f). The side chains of D541, which have previously been identified as critical for Ca<sup>2+</sup> selectivity, permeation and voltage-dependent Mg<sup>2+</sup> block<sup>2</sup>, protrude towards the pore central axis to produce a minimum interatomic distance of 4.6 Å (Fig. 3f–g) at the upper tip of the selectivity filter. Three phenylalanine residues (F530, F533 and F536) in the pore helix, which are conserved in TRPV5–6, but only one of which is conserved in TRPV1–4 (Extended Data Fig. 7), may restrict its dynamics. A relatively static outer pore domain could reflect a key difference between TRPV5/6 and other TRPV channels, which gate in response to various stimuli and thus should display a higher degree of structural plasticity, as exemplified by toxin- and capsaicin-induced conformational changes in TRPV1 (ref. 15).

Below the selectivity filter, the pore widens into a large, mainly hydrophobic cavity (Fig. 3e). Lateral pore portals (Fig. 3a, b) may provide hydrophobic access to this cavity for small molecules or lipids, similar to voltage gated Na<sup>+</sup> channels<sup>19</sup>. The large diameter of the hydrophobic cavity (~13 Å) can easily accommodate a fully hydrated calcium ion, which has an effective diameter of 8–10 Å. The S6 helices cross at the intracellular portion of the channel, where the M577 side chains form the narrow constriction (5.1 Å diameter) and define the lower gate (Figs 3d–f), similar to TRPV2 (ref. 16). Importantly, anomalous diffraction from crystals grown with selenomethionine-labelled protein showed a robust signal (Fig. 3h and Extended Data Fig. 3c), confirming that M577 side chains occlude the pore. Despite high sequence conservation in this region (Extended Data Fig. 7), in TRPV1 the equivalent residue to TRPV6 M577 points away from the pore axis (Extended Data Fig. 6a).

### Cation-binding sites

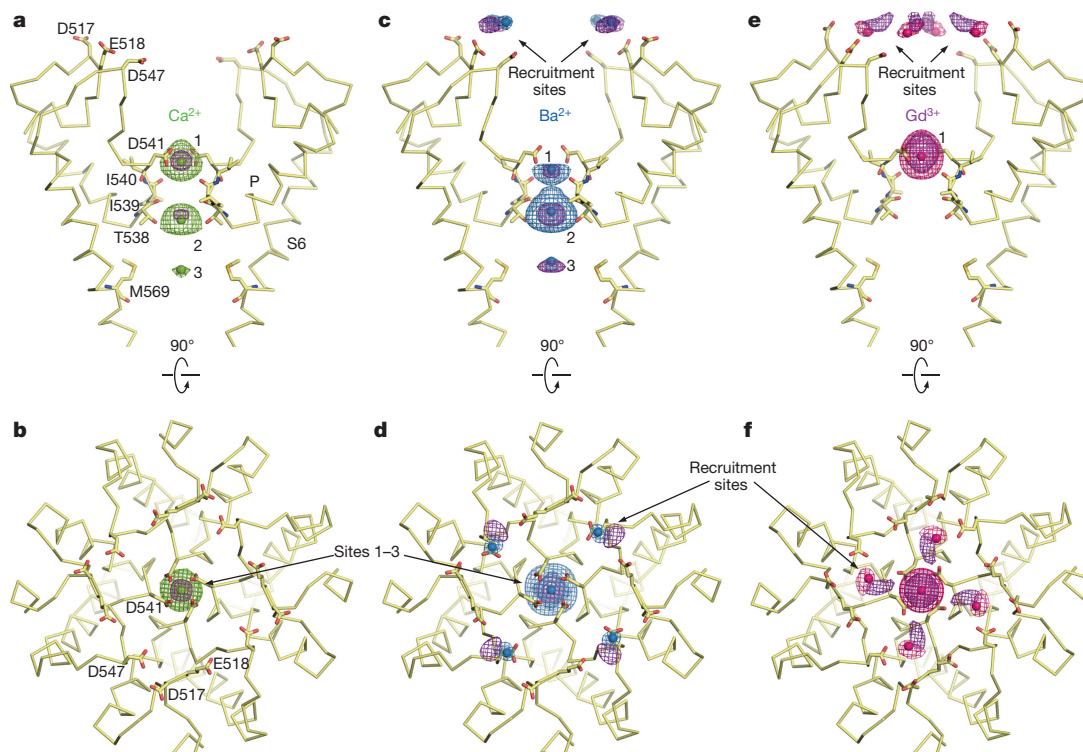
Previous research has resulted in the proposal that TRPV5 and TRPV6 achieve their exceptional Ca<sup>2+</sup> selectivity through binding of Ca<sup>2+</sup> to the selectivity filter<sup>2</sup>. Indeed, we observed a strong  $2F_o - F_c$  density consistent with a bound ion at the central pore axis, surrounded closely by the carbonyl oxygens of D541 side chains (Extended Data Fig. 2e). Since the pore diameter here (4.6 Å, measured between centres of opposing oxygen atoms) is large enough to accommodate a dehydrated calcium ion (typical Ca<sup>2+</sup>–oxygen distance is ~2.4 Å), we contend that the selectivity filter is captured in a Ca<sup>2+</sup>-conducting state. To further resolve cation-binding sites in the pore, we co-crystallized TRPV6<sub>cryst</sub> with Ca<sup>2+</sup>, Ba<sup>2+</sup> or Gd<sup>3+</sup>, which have various permeation and channel-blocking properties<sup>27</sup> (Extended Data Fig. 1) and collected X-ray diffraction data to locate anomalous difference peaks.

The anomalous difference peaks suggest the presence of four types of cation-binding site in the TRPV6<sub>cryst</sub> channel pore (Fig. 4). Notably, two of these (sites 1 and 2) have locations approximately equivalent to Ca<sup>2+</sup> sites in the genetically engineered Ca<sup>2+</sup>-selective channel Ca<sub>v</sub>Ab<sup>28</sup>, but none of them overlap with the putative Ca<sup>2+</sup> site in Ca<sub>v</sub>1.1 (ref. 29) (Extended Data Fig. 6m–o). For Ba<sup>2+</sup> and Gd<sup>3+</sup>, four symmetry-related peaks were observed in the TRPV6<sub>cryst</sub> outer vestibule, in the vicinity of D517, E518 and D547 (Fig. 4c–f). Interestingly, the Ba<sup>2+</sup> and Gd<sup>3+</sup> sites occupy distinct locations, probably because of difference in charge density. Although these signals were not observed for TRPV6<sub>cryst</sub> co-crystallized with Ca<sup>2+</sup> (Fig. 4a–b), presumably because of lower affinity, reduced occupancy or weaker anomalous signal, we speculate that the highly electronegative outer vestibule is involved in the general recruitment of cations towards the extracellular vestibule of



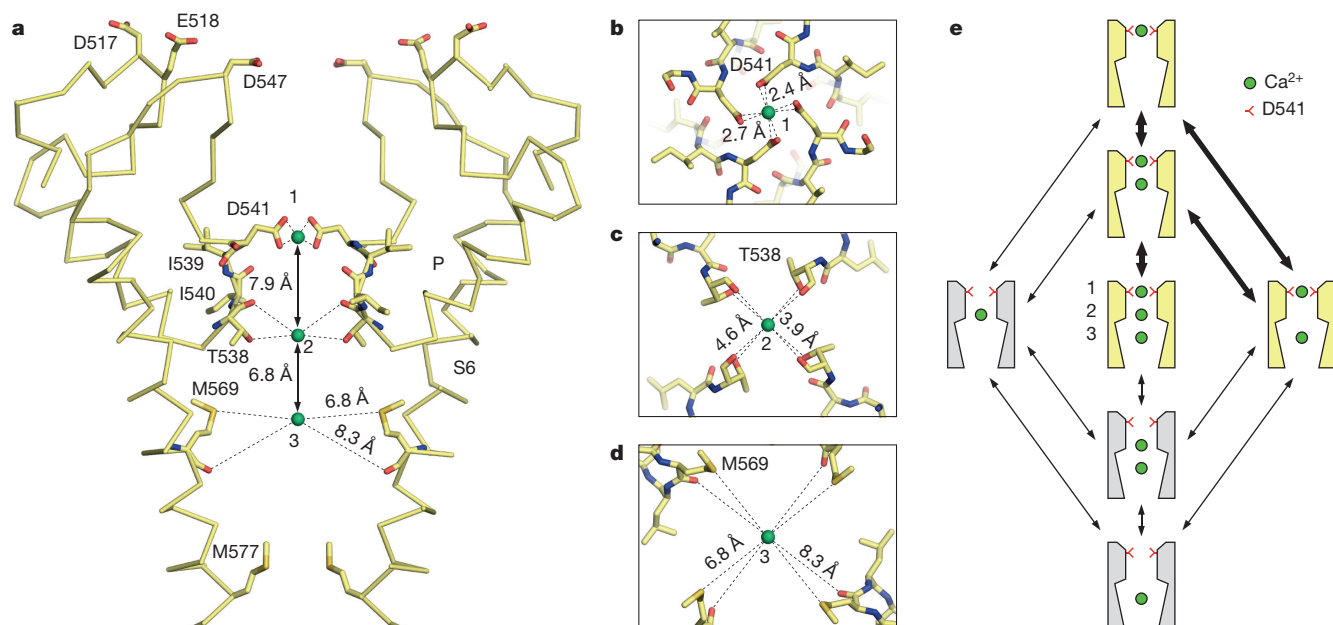
**Figure 3 | Permeation pathway.** **a–c**, Side (**a**), central slice (**b**) and top (**c**) views of TRPV6<sub>cryst</sub> structure in surface representation, coloured by electrostatic potential. **d**, Ribbon diagram of the TRPV6<sub>cryst</sub> tetramer, with ion conduction pathway shown in cyan. **e**, Expanded view of the TRPV6<sub>cryst</sub> pore, with front and back subunits excluded for clarity. Acidic side chains in the extracellular vestibule and pore-lining side chains are shown as sticks. **f**, Radius of the pore calculated using HOLE. D541 and

M577 form narrow constrictions at the selectivity filter and intracellular gate, respectively. **g, h**, Top views of narrow constrictions formed by D541 (**g**) and M577 (**h**). In **h**, blue and pink mesh shows electron density for M577 ( $2F_o - F_c$ , 45–3.25 Å, 1.0 $\sigma$ ) and anomalous difference electron density from selenomethionine-labelled crystal (30–5.00 Å, 3.0 $\sigma$ ), respectively.



**Figure 4 | Cation-binding sites in the TRPV6<sub>cryst</sub> pore.** Side (**a, c, e**) and top (**b, d, f**) views of the TRPV6<sub>cryst</sub> pore, with residues important for cation binding shown in stick representation. Front and back subunits in **a, c** and **e** are removed for clarity. Green, blue and pink mesh shows anomalous difference electron density for Ca<sup>2+</sup> (**a, b**, 38–4.59 Å, 2.7 $\sigma$ ), Ba<sup>2+</sup> (**c, d**, 38–4.59 Å, 3.5 $\sigma$ ) and Gd<sup>3+</sup> (**e, f**, 38–4.59 Å, 7 $\sigma$ ), and ions are shown as spheres of the corresponding colour. Purple mesh shows

simulated-annealing  $F_o - F_c$  electron density maps contoured at 4 $\sigma$  for Ca<sup>2+</sup> (50–3.65 Å), 3 $\sigma$  for Ba<sup>2+</sup> (50–3.85 Å) and 3.5 $\sigma$  for Gd<sup>3+</sup> (50–3.80 Å). The amplitudes of the anomalous peaks are listed in Extended Data Fig. 8c. D547 and E518 side chains are apparently involved in coordination of Ba<sup>2+</sup> ions at the recruitment sites. The Gd<sup>3+</sup> recruitment sites are distinct from Ba<sup>2+</sup> and apparently involve coordination by D517, E518 and D547 side chains.



**Figure 5 | Calcium permeation mechanism.** **a**, Side view of TRPV6<sub>cryst</sub> pore, with front and back subunits removed for clarity. Residues that surround or contribute to cation-binding sites are shown as sticks, and Ca<sup>2+</sup> ions at sites 1, 2 and 3 are shown as green spheres. **b–d**, Top views of Ca<sup>2+</sup> ions at sites 1 (**b**), 2 (**c**) and 3 (**d**), with nearby residues shown as sticks. The interatomic distances illustrated by dashed lines suggest that Ca<sup>2+</sup> is directly coordinated by D541 side chains at site 1, while a hydrated Ca<sup>2+</sup> ion indirectly interacts with the pore at sites 2 and 3. **e**, Schematic

representation of various Ca<sup>2+</sup> occupancy states in TRPV6. Presumed lower energy states are shown in yellow, and most probable transitions are highlighted with bold arrows. Occupancy states in which site 1 is vacant (shown in grey) are likely only to be transiently populated, owing to electrostatic repulsion of D541 side chains. Sufficiently large distances between sites 1, 2 and 3 suggest that electrostatic repulsion between Ca<sup>2+</sup> ions does not preclude simultaneous binding at all three sites.

the TRPV6 channel. Lower affinity of the recruitment sites compared with the main binding site in the centre of the pore for Gd<sup>3+</sup> is consistent with the results of isothermal titration calorimetry experiments (Extended Data Fig. 8a, b).

The strongest anomalous difference peaks for Ca<sup>2+</sup> and Gd<sup>3+</sup> were observed along the central pore axis at or near the same plane as D541 side chains (Fig. 4a, b, e, f), indicating that this locus constitutes the main cation-binding site (site 1). The cation–oxygen distance of 2.4 Å (Fig. 5b) matches the reported average Ca<sup>2+</sup>–oxygen distance calculated from crystal structures of various classes of Ca<sup>2+</sup>-binding proteins<sup>30</sup>. This minimal interatomic distance suggests that the carboxylate oxygens of D541 directly coordinate an at least partly dehydrated Ca<sup>2+</sup> ion at this site. Similarly, structural studies of the hexameric Ca<sup>2+</sup> release-activated channel Orai suggest that Ca<sup>2+</sup> selectivity is achieved by direct coordination of Ca<sup>2+</sup> by a ring of glutamate residues at the extracellular entrance to the pore<sup>31</sup>. By contrast, in Ca<sub>v</sub>Ab, the permeant Ca<sup>2+</sup> ion indirectly interacts with the pore through water molecules<sup>28</sup>. The presence of a robust Gd<sup>3+</sup> signal at site 1 shows that trivalents can bind at D541 as well (Fig. 4e, f).

For Ca<sup>2+</sup> and Ba<sup>2+</sup>, an additional anomalous difference signal is observed at the centre of the pore, 6–8 Å below site 1, between the backbone carbonyls and side-chain hydroxyl groups of T538 (site 2). The greater Ca<sup>2+</sup>/Ba<sup>2+</sup>–oxygen distance at site 2 (~4 Å, Figs 4a, c and 5c) indicates that the cation is equatorially hydrated at this location. Although the chemical environment of site 2 suggests that it binds cations at lower affinity than site 1, the Ba<sup>2+</sup> signal is stronger at this site (Extended Data Fig. 8c). The different relative anomalous peak intensities of sites 1 and 2 for Ca<sup>2+</sup> and Ba<sup>2+</sup>, as well as their slightly different positions at site 1, may arise from the greater size of Ba<sup>2+</sup> (~3 Å diameter) than Ca<sup>2+</sup> (~2 Å diameter). This observation implies that the TRPV6 selectivity filter discriminates ions on the basis of size as well as charge.

Anomalous difference peaks were observed for Ca<sup>2+</sup> and Ba<sup>2+</sup> 6.8 Å below site 2 in the centre of the hydrophobic cavity, at the level of M569 (site 3) (Figs 4a, c and 5a, d). For Ca<sup>2+</sup>, the anomalous peak

at site 3 is less robust (Extended Data Fig. 8c), presumably because of weaker anomalous diffraction properties. The signal at site 3 suggests that cations bound here are ordered by water molecules, which can be held in place by weak hydrogen bonding interactions and pore helix dipoles pointing their partial negative charges towards the centre of the hydrophobic cavity.

### Mechanism of ion permeation

The pore architecture and locations of cation-binding sites in the TRPV6<sub>cryst</sub> structure (Fig. 5a–d) illuminate a potential calcium permeation mechanism (Fig. 5e). The close proximity of carboxylate side chains at site 1 suggests that, in the present pore conformation, the absence of a bound Ca<sup>2+</sup> ion would be energetically unfavourable because of charge repulsion between D541 side chains. Thus, it is likely that a Ca<sup>2+</sup> ion is, in effect, constitutively bound at site 1 and removal of a Ca<sup>2+</sup> ion from site 1 would require immediate replacement with another Ca<sup>2+</sup> ion, necessitating a ‘knock-off’ mechanism of permeation similar to the genetically engineered Ca<sup>2+</sup>-selective channel Ca<sub>v</sub>Ab<sup>28</sup>. Given the large energetic barrier of displacing a Ca<sup>2+</sup> ion at site 1, a substantially high local concentration of Ca<sup>2+</sup> would be necessary for permeation to proceed at physiological membrane voltages. Recruitment sites in the highly electronegative extracellular vestibule might serve this purpose.

As direct coordination by aspartate side chains suggests that site 1 is the highest affinity site for Ca<sup>2+</sup> in TRPV6 channel pore, knock-off from site 1 is likely to be the rate-limiting step for Ca<sup>2+</sup> permeation. After the Ca<sup>2+</sup> ion is knocked off site 1, it moves towards site 2, where it is coordinated through its hydration shell by the backbone carbonyls and sidechain hydroxyls of T538. In Ca<sub>v</sub>Ab<sup>28</sup>, Ca<sup>2+</sup> also binds in the middle of the selectivity filter, at a locus between site 1 and site 2 of TRPV6<sub>cryst</sub> (Extended Data Fig. 6m, n). Although we found no crystallographic evidence for Ca<sup>2+</sup> bound at an equivalent site in TRPV6<sub>cryst</sub>, it is plausible that such a site is occupied transiently during stepwise Ca<sup>2+</sup> permeation. Whether a knock-off is necessary for the Ca<sup>2+</sup> ion to traverse from site 2 to site 3 is unclear, as electrostatic repulsion

between  $\text{Ca}^{2+}$  ions at site 1 and site 2 (and possibly, the aforementioned site between sites 1 and 2) may contribute a driving force. At site 3, the  $\text{Ca}^{2+}$  ion is poised to enter the cell. Since the lower gate is closed in the current TRPV6<sub>cryst</sub> structure, further studies are necessary to elucidate whether its opening affects cation binding in the pore.

Previous observations have suggested that, in addition to  $\text{Ca}^{2+}$ , TRPV6 is permeable to other divalents (with ion permeation sequence  $\text{Ca}^{2+} > \text{Sr}^{2+} \approx \text{Ba}^{2+} > \text{Mn}^{2+}$ )<sup>4</sup> and weakly to trivalents ( $\text{La}^{3+}$  and  $\text{Gd}^{3+}$ )<sup>27</sup> as well. The anomalous difference peaks for  $\text{Ba}^{2+}$  and  $\text{Gd}^{3+}$  indicate that the permeation mechanism of other cations differs from  $\text{Ca}^{2+}$  permeation to varying degrees.  $\text{Ba}^{2+}$ , for example, apparently has a stronger anomalous electron density at site 2 (Extended Data Fig. 8c), which suggests a higher affinity for that site than site 1. Thus, knock-off of  $\text{Ba}^{2+}$  from site 2 to site 3 may be slower and more rate-limiting than knock-off from site 1 to site 2. Larger and more positively charged ions such as  $\text{Gd}^{3+}$  may permeate differently from divalent cations, since their high charge density may preclude simultaneous binding at sites 1 and 2. Nevertheless, trivalents probably block divalents from permeating by virtue of their strong positive charge, which results in higher affinity binding at site 1. Likewise,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  probably block monovalent currents<sup>2</sup> through an analogous mechanism. Further studies will be necessary to elucidate the intricate details of cation permeation and selectivity in epithelial  $\text{Ca}^{2+}$  channels.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 14 January; accepted 6 April 2016.**

**Published online 13 June 2016.**

- Clapham, D. E. TRP channels as cellular sensors. *Nature* **426**, 517–524 (2003).
- Owsianik, G., Talavera, K., Voets, T. & Nilius, B. Permeation and selectivity of TRP channels. *Annu. Rev. Physiol.* **68**, 685–717 (2006).
- Miura, S., Sato, K., Kato-Negishi, M., Teshima, T. & Takeuchi, S. Fluid shear triggers microvilli formation via mechanosensitive activation of TRPV6. *Nature Commun.* **6**, 8871 (2015).
- den Dekker, E., Hoenderop, J. G. J., Nilius, B. & Bindels, R. J. M. The epithelial calcium channels, TRPV5 & TRPV6: from identification towards regulation. *Cell Calcium* **33**, 497–507 (2003).
- Woudenberg-Vrenken, T. E. et al. Functional TRPV6 channels are crucial for transepithelial  $\text{Ca}^{2+}$  absorption. *Am. J. Physiol. Gastrointest. Liver Physiol.* **303**, G879–G885 (2012).
- Bianco, S. D. et al. Marked disturbance of calcium homeostasis in mice with targeted disruption of the *Trpv6* calcium channel gene. *J. Bone Miner. Res.* **22**, 274–285 (2007).
- Weissgerber, P. et al. Male fertility depends on  $\text{Ca}^{2+}$  absorption by TRPV6 in epididymal epithelia. *Sci. Signal.* **4**, ra27 (2011).
- Fecher-Trost, C., Weissgerber, P. & Wissenbach, U. in *Mammalian Transient Receptor Potential (TRP) Cation Channels* (eds Nilius, B. & Flockerzi, V.) Ch. TRPV6 Channels, 359–384 (Springer, 2014).
- Lehen'kyi, V., Raphaël, M. & Prevarskaya, N. The role of the TRPV6 channel in cancer. *J. Physiol. (Lond.)* **590**, 1369–1376 (2012).
- Raphaël, M. et al. TRPV6 calcium channel translocates to the plasma membrane via Orai1-mediated mechanism and controls cancer cell survival. *Proc. Natl Acad. Sci. USA* **111**, E3870–E3879 (2014).
- Bowen, C. V. et al. In vivo detection of human TRPV6-rich tumors with anti-cancer peptides derived from soricidin. *PLoS ONE* **8**, e58866 (2013).
- Bolanz, K. A., Kovacs, G. G., Landowski, C. P. & Hediger, M. A. Tamoxifen inhibits TRPV6 activity via estrogen receptor-independent pathways in TRPV6-expressing MCF-7 breast cancer cells. *Mol. Cancer Res.* **7**, 2000–2010 (2009).
- Hoenderop, J. G. J. et al. Homo- and heterotetrameric architecture of the epithelial  $\text{Ca}^{2+}$  channels TRPV5 and TRPV6. *EMBO J.* **22**, 776–785 (2003).
- Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013).
- Cao, E., Liao, M., Cheng, Y. & Julius, D. TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature* **504**, 113–118 (2013).
- Zubcevic, L. et al. Cryo-electron microscopy structure of the TRPV2 ion channel. *Nature Struct. Mol. Biol.* **23**, 180–186 (2016).
- Paulsen, C. E., Armache, J. P., Gao, Y., Cheng, Y. & Julius, D. Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature* **520**, 511–517 (2015).
- Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent  $\text{K}^{+}$  channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
- Payandeh, J., Scheuer, T., Zheng, N. & Catterall, W. A. The crystal structure of a voltage-gated sodium channel. *Nature* **475**, 353–358 (2011).
- Hite, R. K. et al. Cryo-electron microscopy structure of the Slo2.2  $\text{Na}^{+}$ -activated  $\text{K}^{+}$  channel. *Nature* **527**, 198–203 (2015).
- Lu, P., Boros, S., Chang, Q., Bindels, R. J. & Hoenderop, J. G. The  $\beta$ -glucuronidase klotho exclusively activates the epithelial  $\text{Ca}^{2+}$  channels TRPV5 and TRPV6. *Nephrol. Dial. Transplant.* **23**, 3397–3402 (2008).
- Chang, Q. et al. Molecular determinants in TRPV5 channel assembly. *J. Biol. Chem.* **279**, 54304–54311 (2004).
- de Groot, T. et al. Role of the transient receptor potential vanilloid 5 (TRPV5) protein N terminus in channel activity, tetramerization, and trafficking. *J. Biol. Chem.* **286**, 32132–32139 (2011).
- Cao, C., Zakharian, E., Borbiri, I. & Rohacs, T. Interplay between calmodulin and phosphatidylinositol 4,5-bisphosphate in  $\text{Ca}^{2+}$ -induced inactivation of transient receptor potential vanilloid 6 channels. *J. Biol. Chem.* **288**, 5278–5290 (2013).
- Lambers, T. T., Weidema, A. F., Nilius, B., Hoenderop, J. G. J. & Bindels, R. J. M. Regulation of the mouse epithelial  $\text{Ca}^{2+}$  channel TRPV6 by the  $\text{Ca}^{2+}$ -sensor calmodulin. *J. Biol. Chem.* **279**, 28855–28861 (2004).
- de Groot, T. et al. Molecular mechanisms of calmodulin action on TRPV5 and modulation by parathyroid hormone. *Mol. Cell. Biol.* **31**, 2845–2853 (2011).
- Kovacs, G. G. et al. Heavy metal cations permeate the TRPV6 epithelial cation channel. *Cell Calcium* **49**, 43–55 (2011).
- Tang, L. et al. Structural basis for  $\text{Ca}^{2+}$  selectivity of a voltage-gated calcium channel. *Nature* **505**, 56–61 (2014).
- Wu, J. P. et al. Structure of the voltage-gated calcium channel  $\text{Ca}_v1.1$  complex. *Science* **350**, 1492–1501 (2015).
- Yang, W., Lee, H. W., Hellinga, H. & Yang, J. J. Structural analysis, identification, and design of calcium-binding sites in proteins. *Proteins* **47**, 344–356 (2002).
- Hou, X., Pedi, L., Diver, M. M. & Long, S. B. Crystal structure of the calcium release-activated calcium channel Orai. *Science* **338**, 1308–1313 (2012).

**Acknowledgements** We thank E. C. Twomey and J. M. Sampson for comments on the manuscript, members of the E. C. Greene laboratory for assistance with their fluorimeter, and the Columbia Protein Core facility for assistance with isothermal titration calorimetry (ITC) measurements. We also thank the personnel at beamlines 24-ID-C/E of APS, X25/X29 of NSLS and 5.0.1/5.0.2 of ALS. This work was supported by National Institutes of Health grants R01 NS083660 (A.I.S.) and T32 GM008281 (K.S.), by a Pew Scholar Award in Biomedical Sciences, a Schaefer Research Scholar Award, a Klingenstein Fellowship Award in the Neurosciences and an Irma T. Hirschl Career Scientist Award (A.I.S.).

**Author Contributions** K.S., A.K.S., M.V.Y. and A.I.S. designed the project. K.S., A.K.S. and M.V.Y. performed the experiments. K.S., A.K.S., M.V.Y. and A.I.S. wrote the manuscript.

**Author Information** The structure coordinates have been deposited in Protein Data Bank (PDB) with accession numbers 5IWK, 5IWP, 5IWR and 5IWT for native,  $\text{Ca}^{2+}$ ,  $\text{Ba}^{2+}$  and  $\text{Gd}^{3+}$  data, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.I.S. (as4005@cumc.columbia.edu).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Constructs.** Using fluorescence-detection size-exclusion chromatography (FSEC)<sup>32</sup>, we screened numerous TRPV5 and TRPV6 orthologues fused to enhanced green fluorescent protein (eGFP)<sup>33</sup> at the C terminus and identified rat TRPV6 (GenBank EDM15484.1) as the best candidate for crystallographic trials. The fortuitous spontaneous mutation L495Q generated during gene synthesis was found to increase expression level of rat TRPV6. C-terminal truncation mutants of rat TRPV6-L495Q produced crystals in the C222 space group that diffracted to  $\sim 6$  Å resolution. On the basis of an initial low-resolution molecular replacement solution, we designed numerous mutations aimed at improving crystal packing, including individual substitutions of surface residues, fusions with soluble protein partners and flexible loop deletions. Incorporation of the surface residue mutations L92N and M96Q helped improve the resolution limit of crystals in the C222 space group to  $\sim 4.0$  Å. Further screening of surface residue mutations yielded the amino-acid substitution I62Y, which facilitated crystallization in the P4<sub>2</sub>12 space group and improved diffraction resolution to 3.25 Å. Inspection of protein-mediated crystal contacts in this crystal form shows that cation- $\pi$  and/or hydrogen-bonding interactions involving the side chains of I62Y, K63, K66 and F67 might have permitted crystallization in the P4<sub>2</sub>12 space group and contributed to the improved resolution (Extended Data Fig. 9). The final construct, TRPV6<sub>cryst</sub>, comprises residues 1–668 and contains the point mutations I62Y, L92N, M96Q and L495Q.

**Expression and purification.** TRPV6<sub>cryst</sub> was introduced into a pEG BacMam vector<sup>34</sup> with C-terminal thrombin cleavage site (LVPRG) followed by eGFP and streptavidin affinity tag (WSHPQFEK). Baculovirus was made in Sf9 cells (Thermo Fisher Scientific, mycoplasma test negative). For large-scale expression, suspension-adapted HEK 293S cells lacking *N*-acetyl-glucosaminyltransferase I (GnTI) (ATCC, mycoplasma test negative) were grown in Freestyle 293 media (Life Technologies) supplemented with 2% FBS at 37°C in the presence of 5% CO<sub>2</sub>. The culture was transduced with P2 baculovirus once cells reached a density of  $2.5 \times 10^6$  to  $3.5 \times 10^6$  per millilitre. After 8–12 h, 10 mM sodium butyrate was added and the temperature was changed to 30°C. Cells were harvested 48–72 h after transduction and resuspended in a buffer containing 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM  $\beta$ -mercaptoethanol ( $\beta$ ME), 0.8  $\mu$ M aprotinin, 2  $\mu$ g ml<sup>-1</sup> leupeptin, 2 mM pepstatin A and 1 mM phenylmethylsulfonyl fluoride (PMSF). The cells were disrupted using a Misonix Sonicator (12  $\times$  15 s, power level 7), and the resulting homogenate was clarified using a Sorval centrifuge at 9,900g for 15 min. Crude membranes were collected by ultracentrifugation for 1 h in a Beckman Ti45 rotor at 186,000g. The membranes were mechanically homogenized and subsequently solubilized for 2–4 h in a buffer containing 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM  $\beta$ ME, 20 mM *n*-dodecyl- $\beta$ -D-maltopyranoside (DDM), 0.8  $\mu$ M aprotinin, 2  $\mu$ g ml<sup>-1</sup> leupeptin, 2 mM pepstatin A and 1 mM PMSF. After insoluble material was removed by ultracentrifugation, streptavidin-linked resin was added to the supernatant and rotated for 4–16 h. Resin was washed with 10 column volumes of wash buffer containing 150 mM NaCl, 20 mM Tris pH 8.0, 1 mM  $\beta$ ME and 1 mM DDM, and the protein was eluted using wash buffer supplemented with 2.5 mM D-desthiobiotin. The eluted fusion protein was concentrated to  $\sim 1.0$  mg ml<sup>-1</sup> and digested with thrombin at a mass ratio of 1:100 (thrombin:protein) for 1.5 h at 22°C. The digested protein was concentrated and injected into a Superose 6 column equilibrated in a buffer composed of 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM  $\beta$ ME and 0.5 mM DDM. Tris(2-carboxyethyl)phosphine (TCEP; 10 mM) was added to fractions with elution time corresponding to the tetrameric channel, and protein was concentrated to 2.5–3.0 mg ml<sup>-1</sup> using a 100 kDa MWCO concentrator. All purification steps were conducted on ice or at 4°C. Typical purifications yielded  $\sim 1$  mg of purified protein per litre of transduced cells.

Protocols to express selenomethionine-labelled protein in HEK cells were adapted from literature<sup>35</sup>. Six to 8 h after transduction, cells were pelleted and resuspended in DMEM (Life Technologies) supplemented with 10% FBS and lacking L-methionine. After shaking methionine-depleted cells for 6 h at 37°C, 60 mg of L-selenomethionine was added per litre of cells. Thirty-six to 48 h after transduction, cells were harvested and protein was purified using the same protocol as described above, except for the addition of 4 mM L-methionine to all purification buffers, excluding the final gel filtration buffer. This procedure yielded  $\sim 0.4$  mg of selenomethionine-labelled protein per litre of transduced cells.

**Crystallization and structure determination.** Initial high-throughput vapour diffusion crystallization screens showed that purified TRPV6<sub>cryst</sub> crystallizes in numerous conditions containing low molecular mass polyethylene glycols (PEG 300, PEG 350 monomethyl ether (MME), PEG 400 or PEG 550 MME). The best crystals were grown using a reservoir solution consisting of 20–24% PEG 350

MME, 100 mM NaCl and 100 mM Tris-HCl pH 8.0–8.5. To increase crystal size, 50 mM ammonium formate was added to the protein immediately before crystallization. Two microlitres of protein were mixed with 1.0–1.2  $\mu$ l of reservoir solution, and incubated at 20°C in hanging-drop vapour diffusion trays. Crystals grew as thin plates and reached full size ( $\sim 400 \mu\text{m} \times \sim 120 \mu\text{m} \times \sim 20 \mu\text{m}$ ) within 2 weeks. Crystals were cryoprotected by incubating for a short time in a solution containing 33–36% PEG 350 MME, 100 mM NaCl, 100 mM Tris-HCl pH 8.2, 0.5 mM DDM and 50 mM ammonium formate, and flash frozen in liquid nitrogen. To obtain crystals with Ca<sup>2+</sup>, Ba<sup>2+</sup> or Gd<sup>3+</sup>, protein was incubated with 10 mM CaCl<sub>2</sub>, 10 mM BaCl<sub>2</sub> or 1 mM GdCl<sub>3</sub>, respectively, for at least 1 h at 4°C before crystallization. Crystals of selenomethionine-labelled protein were grown and cryoprotected using the same procedure as crystals of native protein.

Diffraction data collected at APS (beamlines 24-ID-C/E), NSLS (beamlines X25 or X29) or ALS (beamlines 5.0.1 or 5.0.2) were processed using XDS<sup>36</sup> or HKL2000 (ref. 37). The initial structural solution was obtained by molecular replacement using Phaser<sup>38</sup> and the structure of mouse TRPV6 ankyrin domain (PDB accession number 2RFA)<sup>39</sup> as a search probe and the rest of the molecule was iteratively built using rat TRPV1 structure (PDB accession number 3J5P)<sup>14</sup> as a guide. The model encompasses most of the polypeptide (residues 27–637), excluding parts of the S2–S3 linker (residues 409–416) and S4–S5 linker (residues 471–479), which were not clearly visible in the electron density map. The model was refined by alternating cycles of building in COOT<sup>40</sup> and automatic refinement in Phenix<sup>41</sup> or Refmac<sup>42</sup>. Correct sequence registry was aided by anomalous difference Fourier maps calculated from crystals grown in the presence of 10 mM Ca<sup>2+</sup> to highlight sulfur atoms of cysteines and methionines, and from crystals labelled with selenomethionine to highlight selenium atoms (Extended Data Fig. 3). To confirm sequence registry in the C-terminal region, where native methionines are absent, selenomethionine-labelled crystals were produced for protein containing a methionine substitution at L630 (L630M). The anomalous difference Fourier maps were calculated from X-ray diffraction data collected at 1.75 Å for Ca<sup>2+</sup> and Ba<sup>2+</sup>, 1.56 Å for Gd<sup>3+</sup> and 0.979 Å for selenium. All structural figures were prepared in PyMol<sup>43</sup>. Surface representation of the ion permeation pathway was generated using the PyMol plugin version of Caver<sup>44</sup>. The pore radius was calculated using HOLE<sup>45</sup>.

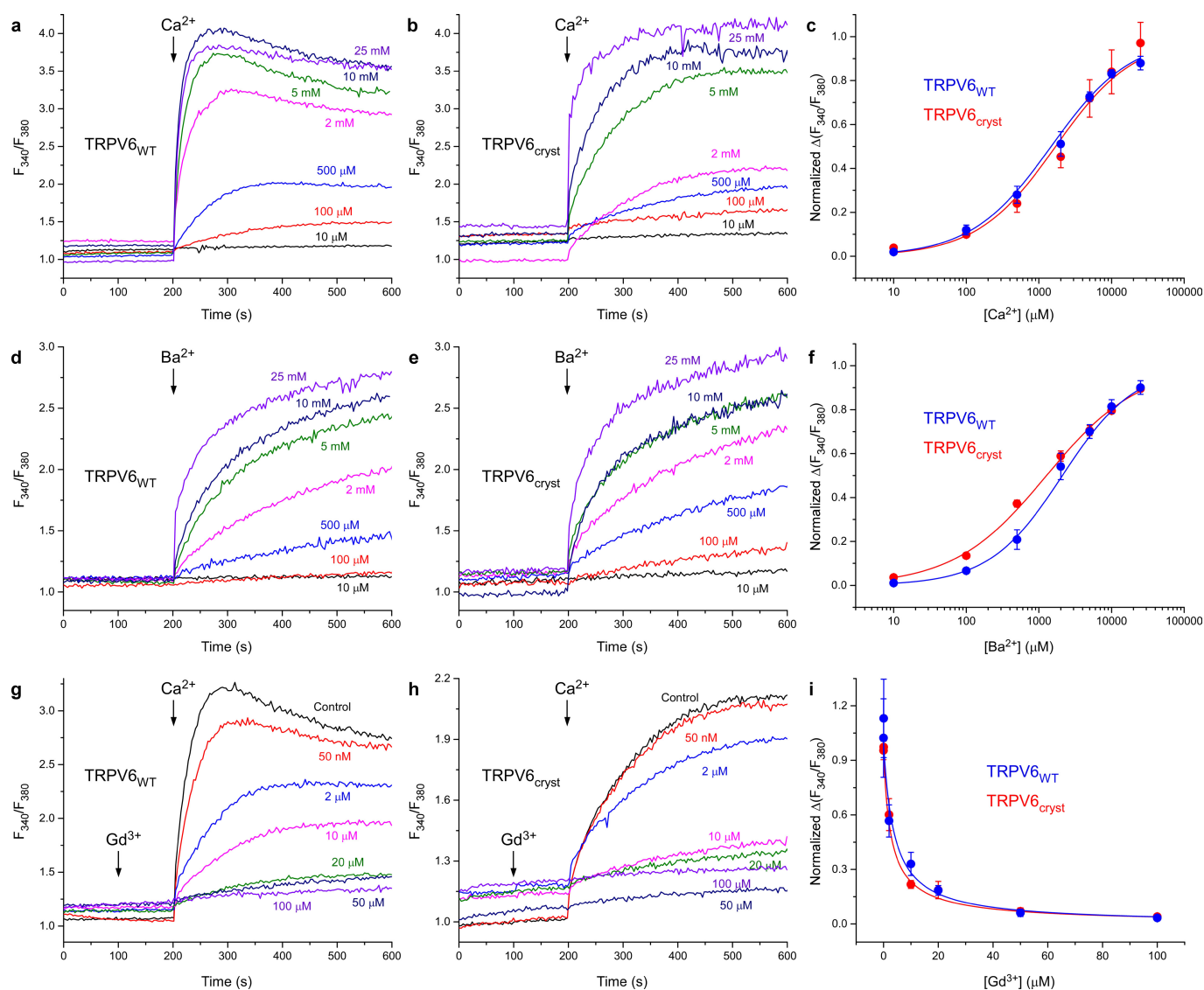
**Fura-2 AM measurements.** Wild-type rat TRPV6 or TRPV6<sub>cryst</sub> fused to C-terminal strep tag was expressed in HEK cells as described above. Forty-eight to 72 h after transduction, cells were harvested by centrifugation at 600g for 5 min. The cells were resuspended in pre-warmed modified HBS (118 mM NaCl, 4.8 mM KCl, 1 mM MgCl<sub>2</sub>, 5 mM D-glucose, 10 mM HEPES pH 7.4) containing 5  $\mu$ g ml<sup>-1</sup> of Fura-2 AM (Life Technologies) and incubated at 37°C for 45 min. The loaded cells were then centrifuged for 5 min at 600g, and resuspended in pre-warmed, modified HBS and incubated again at 37°C for 20–30 min in the dark. The cells were subsequently pelleted and washed twice, then resuspended in modified HBS for experiments. The cells were kept on ice in the dark for maximum of  $\sim 2$  h before fluorescence measurements, which were conducted using a QuantaMaster 40 spectrofluorometer (Photon Technology International) at  $\sim 25^\circ\text{C}$  in a quartz cuvette under constant stirring. Intracellular Ca<sup>2+</sup> was measured by taking the ratio of two excitation wavelengths (340 and 380 nm) at one emission wavelength (510 nm). The excitation wavelength was switched at 1-s intervals.

**Isothermal titration calorimetry experiments.** To study the energetics of Gd<sup>3+</sup> binding, we performed ITC experiments. For these, we used a MicroCal Auto-iTC200 (Malvern Instruments) instrument at the Columbia University ITC Facility. Wild-type TRPV6 protein was purified in buffer containing 20 mM Tris, 150 mM NaCl, 1 mM DDM and 1 mM  $\beta$ ME (buffer A), and the same buffer A was also used to dissolve the desired concentrations of Gd<sup>3+</sup> to avoid buffer mismatch. The experiments were performed at 25°C using 2- $\mu$ l volume injections for the titration and 700 rpm stirring speed for mixing the reactants. The experiments were performed by titrating 700  $\mu$ M Gd<sup>3+</sup> (by robotically controlled syringe) to 6.38- $\mu$ M TRPV6 (in cell) at 3-min intervals. The control experiments were performed to calculate the heat of dilution for each injection by injecting the same volumes of Gd<sup>3+</sup> into buffer A. The data were analysed using a specialized program in Origin (MicroCal ITC).

**Cysteine crosslinking experiments.** For SDS-PAGE and FSEC analysis, cysteine substitutions were introduced into the TRPV6<sub>cryst</sub> background with five exposed cysteines mutated to alanine or serine (C14S, C20S, C70A, C610A and C618A), and the surface mutation I62Y was reverted to the native isoleucine. Cysteine-substituted mutants with C-terminal eGFP and streptavidin affinity tag were expressed in HEK cells in the same way as protein for crystallization and purified with a modified protocol. Crude cell pellets were resuspended in buffer containing 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM  $\beta$ ME, 20 mM DDM, 0.8  $\mu$ M aprotinin, 2  $\mu$ g ml<sup>-1</sup> leupeptin, 2 mM pepstatin A, 1 mM PMSF and stirred for 1–3 h. After insoluble material was removed by ultracentrifugation, streptavidin-linked resin was added to the supernatant and rotated for 4–16 h. Further steps were

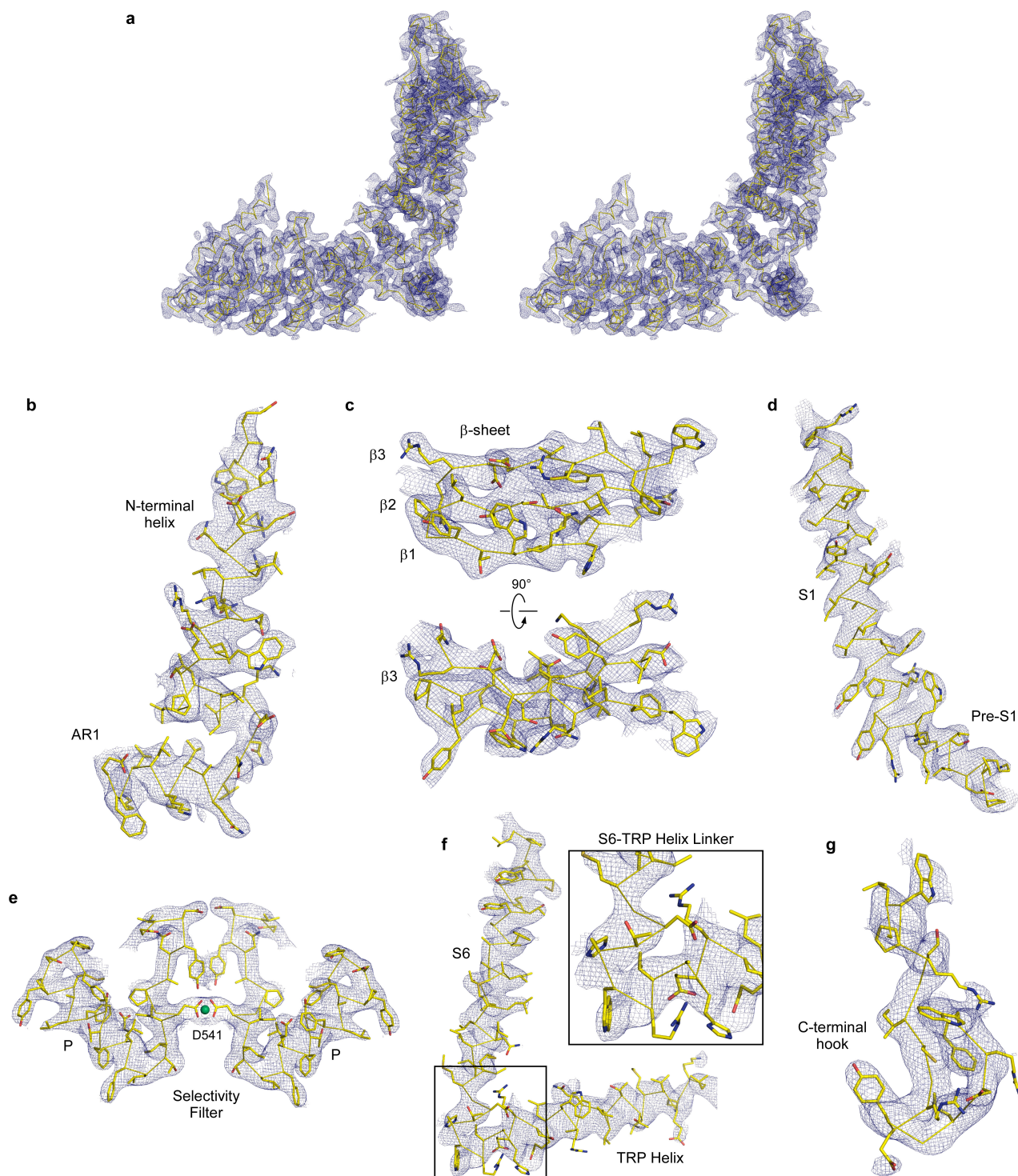
performed in an identical manner to protein purification for crystallization as described above, with the exceptions that the final gel filtration buffer lacked  $\beta$ ME, and TCEP was not added to purified protein. Within 24 h of purification, the protein samples were run on a 4–20% SDS–PAGE and visualized by Coomassie blue staining. A small portion of protein was subjected to FSEC analysis<sup>32</sup>.

32. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
33. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
34. Goehring, A. *et al.* Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nature Protocols* **9**, 2574–2585 (2014).
35. Barton, W. A., Tzvetkova-Robev, D., Erdjument-Bromage, H., Tempst, P. & Nikolov, D. B. Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. *Protein Sci.* **15**, 2008–2013 (2006).
36. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
37. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
38. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32–41 (2007).
39. Phelps, C. B., Huang, R. J., Lishko, P. V., Wang, R. R. & Gaudet, R. Structural analyses of the ankyrin repeat domain of TRPV6 and related TRPV ion channels. *Biochemistry* **47**, 2476–2484 (2008).
40. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
41. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
42. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
43. DeLano, W. L. *The PyMol Molecular Graphics System* (DeLano Scientific, 2002).
44. Chovancova, E. *et al.* CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **8**, e1002708 (2012).
45. Smart, O. S., Neduvellil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360, 376 (1996).
46. Derler, I. *et al.* Dynamic but not constitutive association of calmodulin with rat TRPV6 channels enables fine tuning of  $\text{Ca}^{2+}$ -dependent inactivation. *J. Physiol. (Lond.)* **577**, 31–44 (2006).
47. Lishko, P. V., Procko, E., Jin, X., Phelps, C. B. & Gaudet, R. The ankyrin repeats of TRPV1 bind multiple ligands and modulate channel sensitivity. *Neuron* **54**, 905–918 (2007).
48. Phelps, C. B., Wang, R. R., Choo, S. S. & Gaudet, R. Differential regulation of TRPV1, TRPV3, and TRPV4 sensitivity through a conserved binding site on the ankyrin repeat domain. *J. Biol. Chem.* **285**, 731–740 (2010).



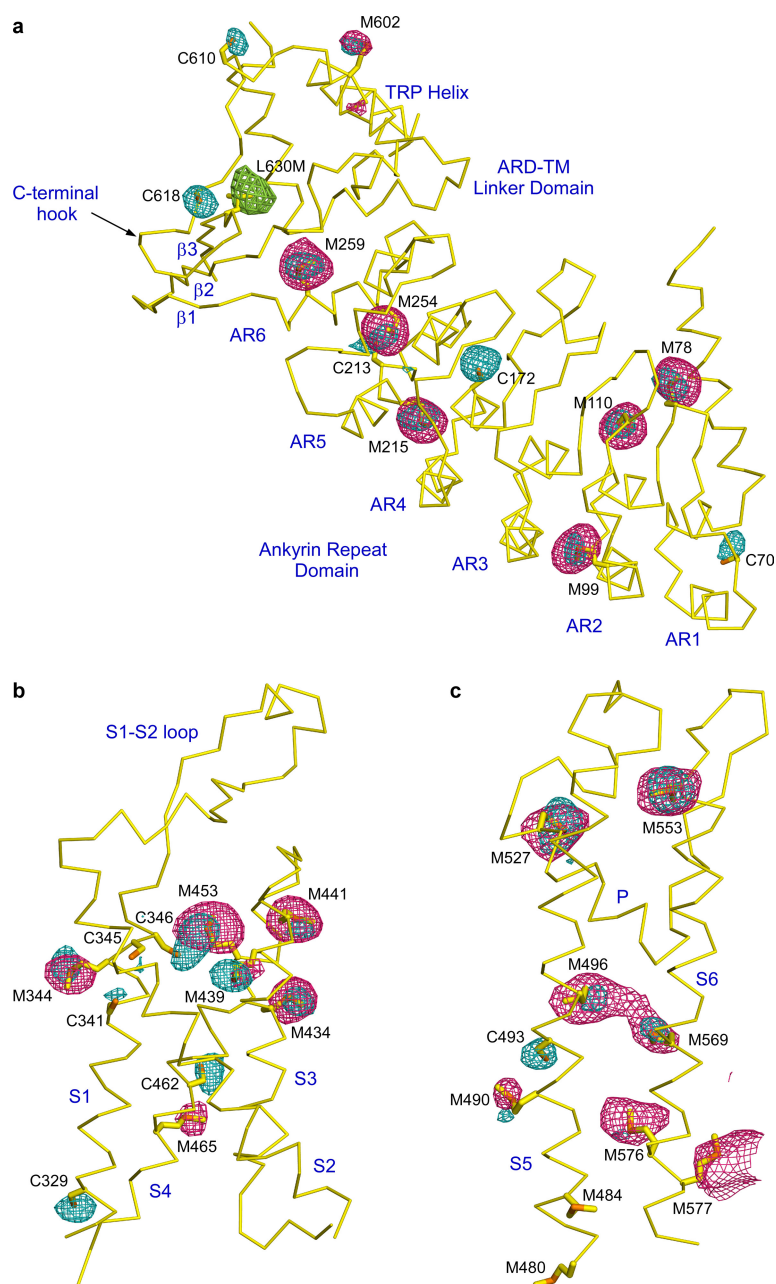
**Extended Data Figure 1 | Functional characterization of wild-type rat TRPV6 and TRPV6<sub>cryst</sub>.** **a, b, d, e, g, h,** Representative ratiometric fluorescence measurements for HEK cells expressing wild-type rat TRPV6 (**a, d, g**) or TRPV6<sub>cryst</sub> (**b, e, h**). Arrows indicate the time at which the corresponding ion was added. After resuspending the cells in nominally calcium-free buffer, addition of Ca<sup>2+</sup> (**a, b**) or Ba<sup>2+</sup> (**d, e**) resulted in robust concentration-dependent increase in Fura-2 signal for both wild-type rat TRPV6 and TRPV6<sub>cryst</sub>. In contrast, pre-incubation of cells with increasing concentrations of Gd<sup>3+</sup> resulted in concentration-dependent reduction in Fura-2 signal for both wild type (**g**) and TRPV6<sub>cryst</sub> (**h**), consistent with Gd<sup>3+</sup> inhibition of wild-type TRPV6 demonstrated previously using <sup>45</sup>Ca<sup>2+</sup> uptake measurements<sup>27</sup>. **c–f,** Dose-response curves for Ca<sup>2+</sup> (**c**) and Ba<sup>2+</sup> (**f**) permeation calculated for wild type (blue) and TRPV6<sub>cryst</sub> (red) ( $n = 3$  for all measurements). The changes in the fluorescence intensity ratio at 340 and 380 nm ( $F_{340}/F_{380}$ ) were normalized to their approximated maximal values at saturating concentrations of Ca<sup>2+</sup> or Ba<sup>2+</sup>, respectively. The apparent values of

half-maximum effective concentration ( $EC_{50}$ ) for TRPV6<sub>cryst</sub> ( $1.70 \pm 0.26$  mM for Ca<sup>2+</sup> and  $1.27 \pm 0.67$  mM for Ba<sup>2+</sup>) are similar to wild type ( $1.47 \pm 0.80$  mM for Ca<sup>2+</sup> and  $1.91 \pm 0.74$  mM for Ba<sup>2+</sup>). **i,** Dose-response curves for Gd<sup>3+</sup> inhibition calculated for wild type (blue) and TRPV6<sub>cryst</sub> (red) ( $n = 3$  for all measurements). The changes  $F_{340}/F_{380}$  evoked by addition of 2 mM Ca<sup>2+</sup> after pre-incubation with various concentrations of Gd<sup>3+</sup> were normalized to the maximal change in  $F_{340}/F_{380}$  after addition of 2 mM Ca<sup>2+</sup> in the absence of Gd<sup>3+</sup>. The apparent values of half-maximum inhibitory concentration ( $IC_{50}$ ) for wild type ( $3.87 \pm 0.83$  μM) are comparable to TRPV6<sub>cryst</sub> ( $2.57 \pm 0.28$  μM). Overall, the mutations introduced to crystallize TRPV6 did not significantly alter its cation permeation and inhibition properties. The absence of time-dependent decay of the Fura-2 AM signal in the case of TRPV6<sub>cryst</sub> is presumably due to its C-terminal truncation, which eliminated a calmodulin-binding site involved in Ca<sup>2+</sup>-dependent inactivation of TRPV6 (ref. 46). Error bars, s.e.m.



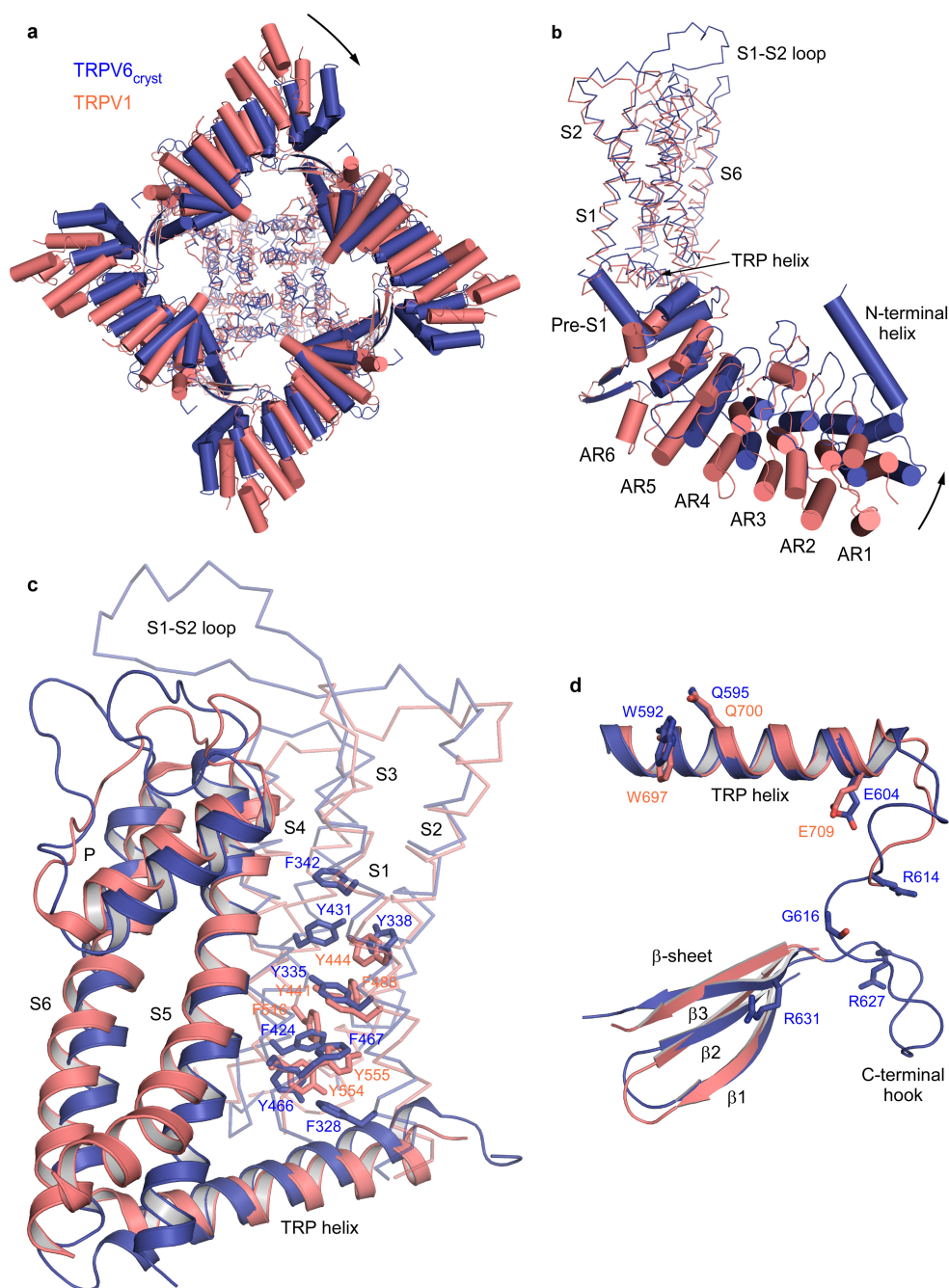
**Extended Data Figure 2 | Electron density.** **a**, Stereo view of  $2F_o - F_c$  electron density map (blue mesh, 45–3.25 Å,  $1.0\sigma$ ) superimposed onto a ribbon model for the entire TRPV6<sub>cryst</sub> monomer. **b–g**, Close-up views of the  $2F_o - F_c$  map for various portions of TRPV6<sub>cryst</sub> model, with side chains shown in stick representation. In **e**, two diagonally opposed

subunits are shown to clarify the position of the central pore axis, and the bound  $\text{Ca}^{2+}$  ion is shown as a green sphere. In **f**, inset shows expanded view of the boxed region, demonstrating electron density for connectivity in the S6-TRP helix linker that is distinct from other TRP channel structures<sup>14,16,17</sup>.



**Extended Data Figure 3 | Anomalous difference Fourier maps for sulfur and selenium.** a–c, Fragments of the TRPV6<sub>cryst</sub> model (yellow ribbon) superimposed onto anomalous difference Fourier maps from X-ray diffraction data collected at 1.75 Å from crystals grown in 10 mM Ca<sup>2+</sup> (cyan mesh, 38–4.59 Å, 3.0 $\sigma$ ) and at 0.979 Å from selenomethionine-labelled crystals (pink mesh, 30–5.00 Å, 3.2 $\sigma$ ) of TRPV6<sub>cryst</sub>. Anomalous signal collected from a selenomethionine-labelled crystal of TRPV6<sub>cryst</sub>

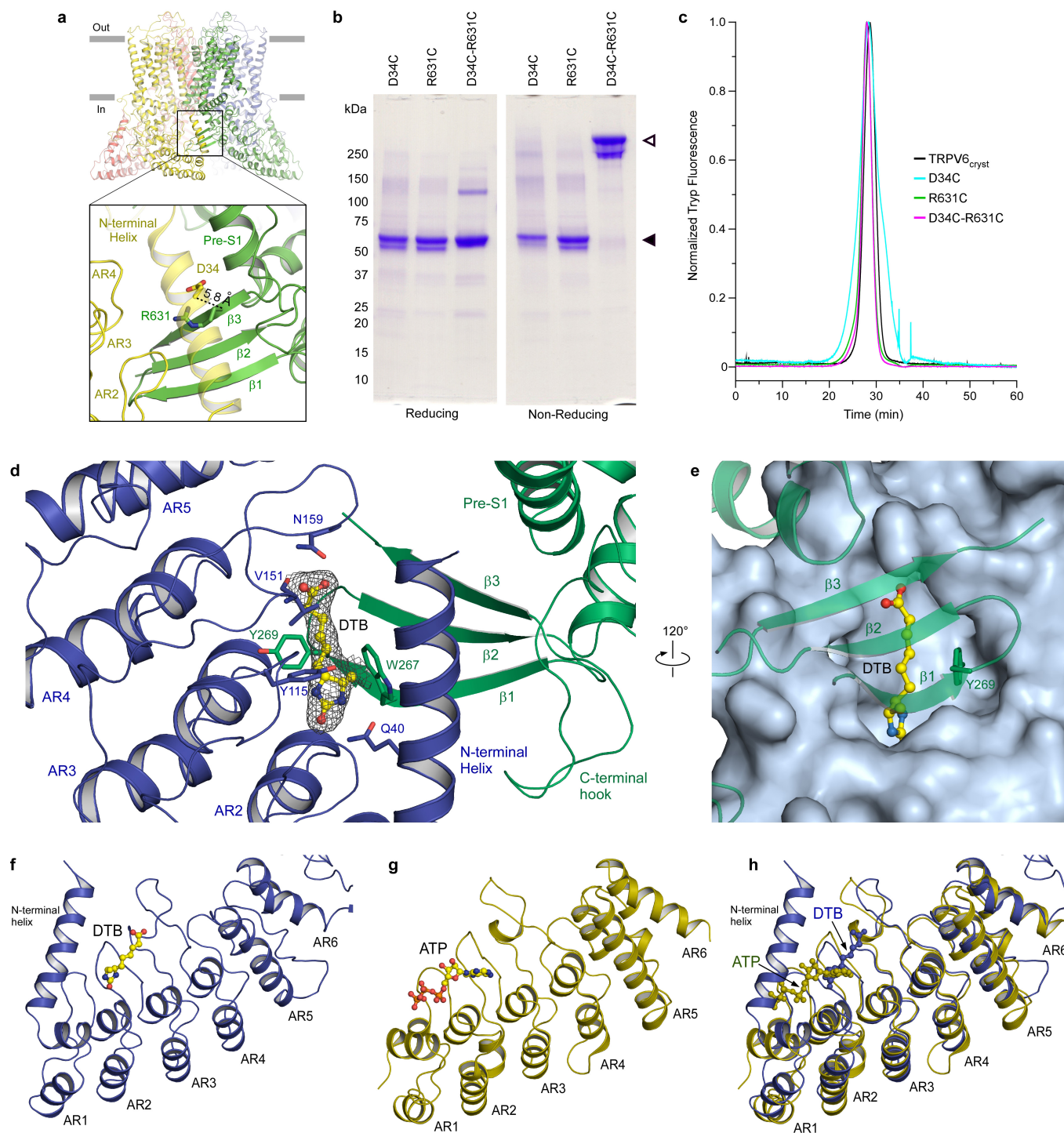
with L630M substitution (a, green mesh, 30–7.20 Å, 3.2 $\sigma$ ) was used to aid registry in the C-terminal  $\beta$ 3 strand. Domains are labelled in blue. Cysteine and methionine residues are shown as sticks and labelled. Sulfur anomalous difference peaks were observed for all cysteines in the TRPV6<sub>cryst</sub> model. Selenium anomalous difference peaks were observed for all methionines in the model, except for M480 and M484 in S5, presumably because of flexibility.



#### Extended Data Figure 4 | Comparison of TRPV6<sub>cryst</sub> and TRPV1.

**a**, Bottom-up view of TRPV6<sub>cryst</sub> (blue) and TRPV1 (salmon) tetramers, with ankyrin repeat domain and linker domain helices shown as cylinders. When S1–S4 domains are aligned, as shown, the cytoplasmic skirt of TRPV6 is rotated clockwise with respect to the cytoplasmic skirt of TRPV1. **b**, Side view of TRPV6<sub>cryst</sub> (blue) and TRPV1 (salmon) monomers with S1–S4 domain based alignment. The ankyrin repeat domain of TRPV1 extends slightly further into the cytoplasm than TRPV6<sub>cryst</sub>. **c**, Alignment of TRPV6<sub>cryst</sub> (blue) and TRPV1 (salmon)

transmembrane domains. Adjacent S1–S4 and pore domains are shown for comparison. Similar to TRPV1, aromatic residues pack against each other to immobilize the TRPV6<sub>cryst</sub> S1–S4 domain core (shown as sticks). The absence of curvature in S5 and the long extracellular S1–S2 loop protruding towards the pore are distinct features of the TRPV6<sub>cryst</sub> transmembrane domain. **d**, Alignment of the TRPV6<sub>cryst</sub> TRP helix, C-terminal hook and three stranded  $\beta$ -sheet with homologous domains in the TRPV1. Conserved residues (Extended Data Fig. 7) are shown in stick representation.

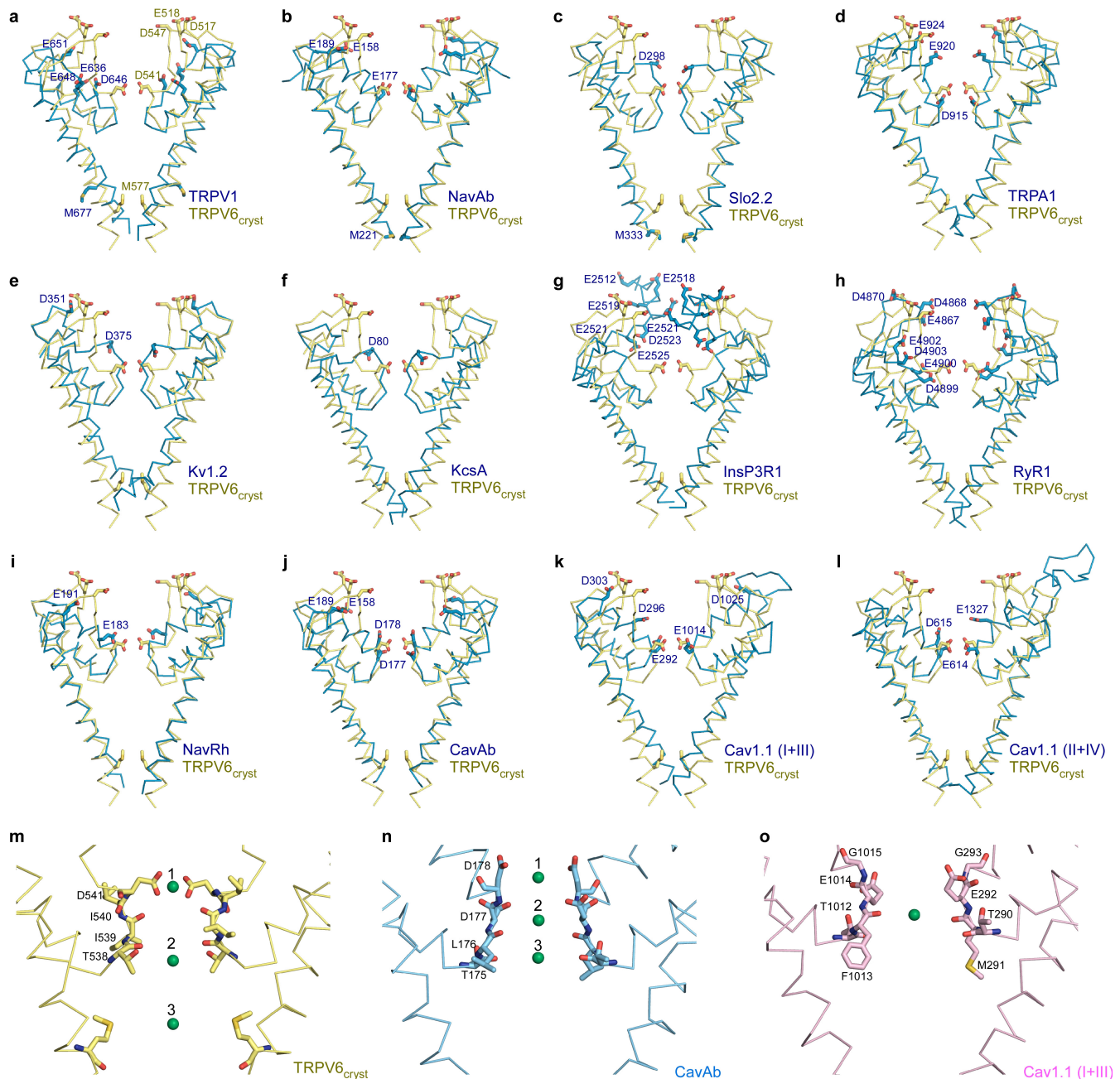


Extended Data Figure 5 | See next page for caption.

**Extended Data Figure 5 | Cysteine crosslinking at the intracellular skirt interface and putative desthiobiotin-binding site at the intracellular intersubunit interface.**

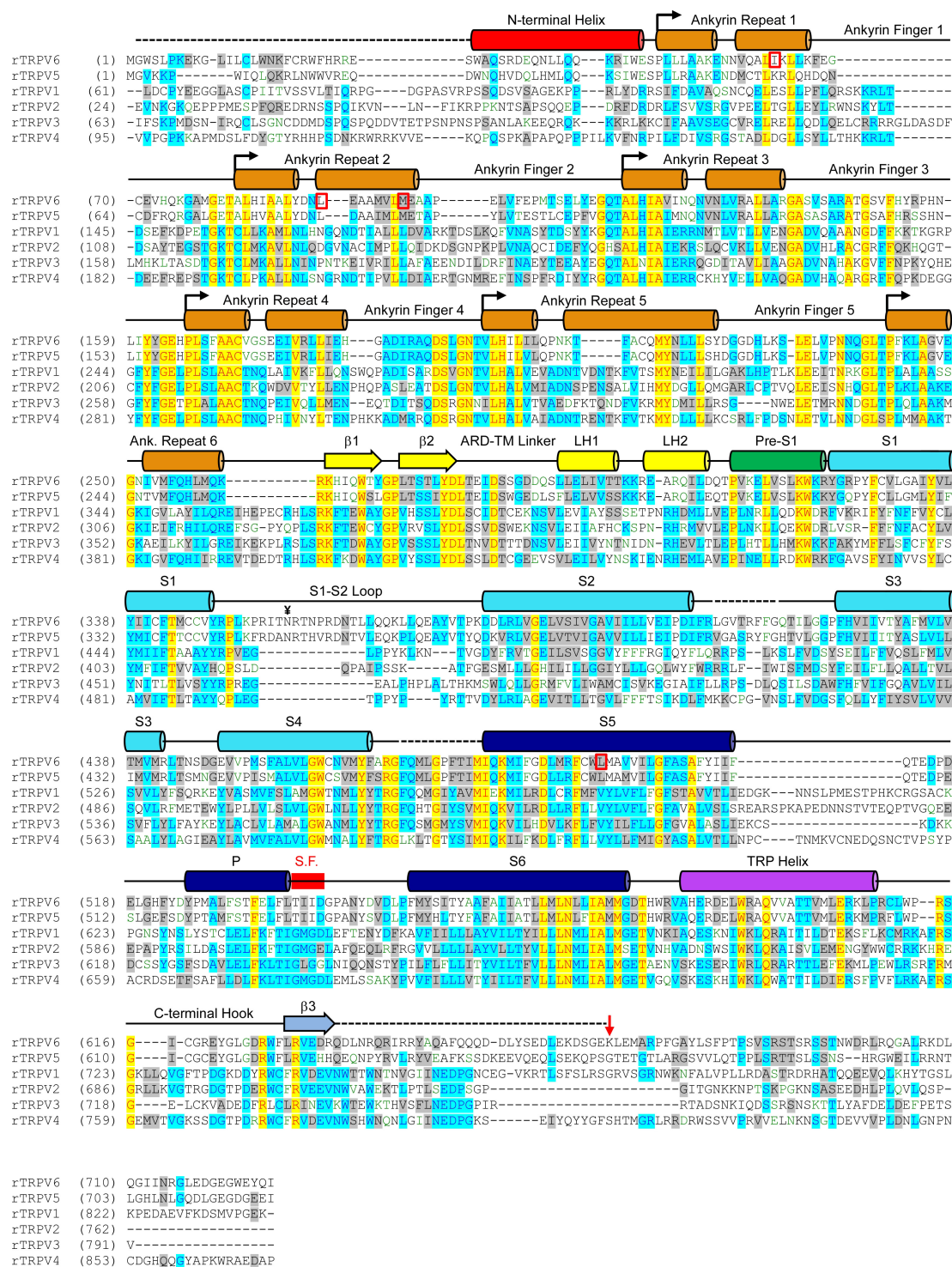
**a**, The TRPV6<sub>cryst</sub> tetramer with each subunit coloured differently (top) and expanded view of boxed region (bottom), with cysteine-substituted residues shown as sticks. Dashed line and label show C<sub>α</sub>–C<sub>α</sub> distance. **b**, SDS–PAGE (4–20% gradient gel) analysis of purified TRPV6 cysteine-substituted mutants in the presence (left) and absence (right) of reducing agent. Cysteines were introduced into a background construct (TRPV6<sub>CysKO</sub>), in which exposed cysteines in TRPV6<sub>cryst</sub> were mutated to serine or alanine (C14S, C20S, C70A, C610A and C618A) to prevent non-specific aggregation. Positions corresponding to monomer and tetramer bands are indicated by filled and open triangles, respectively. The appearance of a robust band corresponding to covalently crosslinked tetramer in the D34C–R631C double mutant indicates that the interacting N-terminal helix (which precedes the S1–S4 domain) and β3 strand (which follows the TRP helix) are from different protomers. Taken together with the S6-TRP helix linker connectivity (Extended Data Fig. 2f) that is different from TRPV1/2 (refs 14, 16) and TRPA1 (ref. 17), these data suggest a non-swapped arrangement of the pore and S1–S4 domains; if the canonical domain-swapped arrangement were true, the interacting N-terminal helix and β3 strand would be from the same monomer and no crosslinked high molecular mass species would form. However, in the absence of interpretable density for the S4–S5 linker, we suggest cautious

interpretation of this domain arrangement. **c**, FSEC analysis of purified TRPV6<sub>CysKO</sub> crosslink mutants in the absence of reducing agent. Each trace shows a single major peak with elution time corresponding to the TRPV6<sub>cryst</sub> tetramer (black trace). **d**, **e**, The putative DTB-binding site is composed of a pocket formed by the N-terminal helix and ankyrin repeats 2–4 of one subunit (blue) and the linker domain of an adjacent subunit (green). DTB is shown as ball and stick, with 2F<sub>o</sub> – F<sub>c</sub> density shown as grey mesh (45–3.25 Å, 1.0σ). In **d**, residues that contact DTB are shown as sticks. In **e**, the binding pocket is shown in surface representation. Interestingly, the DTB-binding site overlaps with the ATP-binding site revealed in the ankyrin domain crystal structure of TRPV1 (ref. 47), which was later demonstrated to be conserved in TRPV3 and TRPV4 (ref. 48). The presence of DTB close to this location in TRPV6 corroborates the assertion made in ref. 14 that ligands bound in this region modulate activity by perturbing subunit interactions. Further work is necessary to establish a functional role, if any, of DTB-like compounds on TRPV6 function. **f–h**, Comparison of the putative DTB-binding site in TRPV6<sub>cryst</sub> (**f**) and the ATP-binding site in the crystal structure of the TRPV1 ankyrin domain (**g**, PDB accession number 2PNN). DTB and ATP are shown in ball and stick. While the ATP-binding site in TRPV1 is shifted towards ankyrin repeat finger 1, both binding sites are located at intersubunit interfaces, as illustrated when the structures are superimposed (**h**).



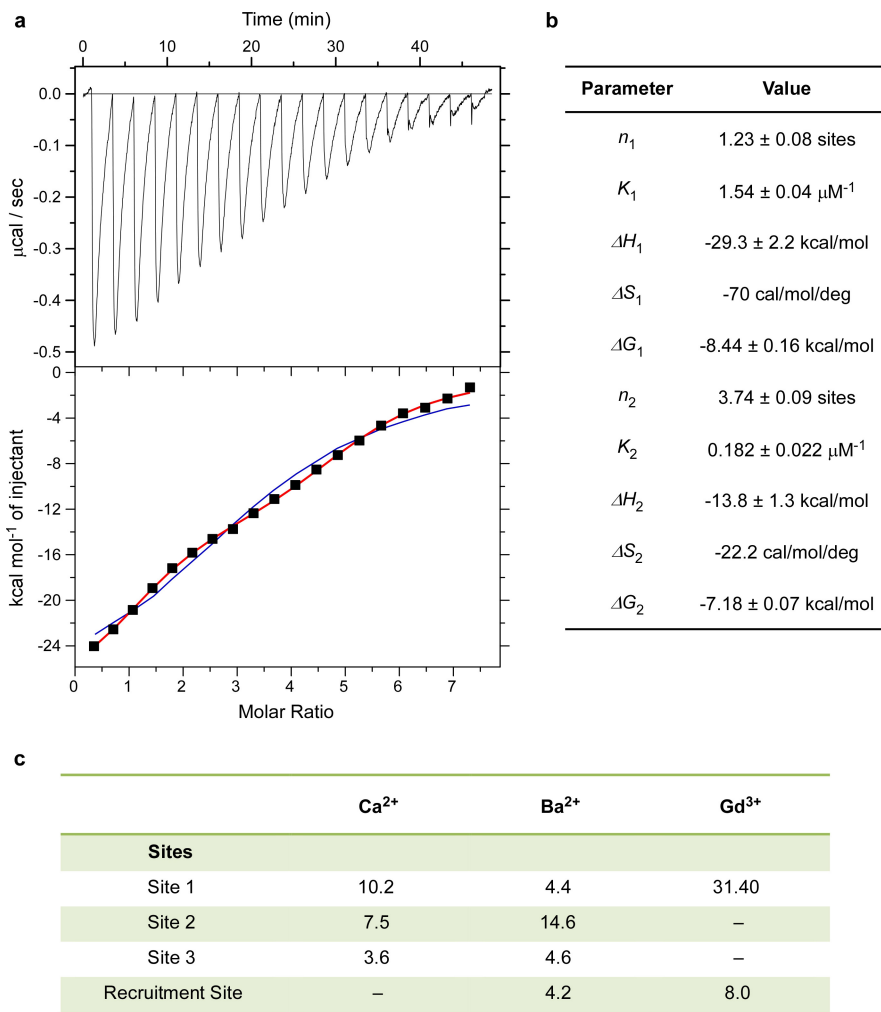
**Extended Data Figure 6 | Comparison of the ion channel pore in TRPV6<sub>cryst</sub> with other tetrameric ion channels.** **a–l**, The pore of TRPV6<sub>cryst</sub> (yellow ribbon) was aligned with TRPV1 (**a**, PDB accession number 3J5P), NavAb (**b**, PDB accession number 3RVY), Slo2.2 (**c**, PDB accession number 5A6E), TRPA1 (**d**, PDB accession number 3J9P), Kv1.2 (**e**, PDB accession number 2R9R), KcsA (**f**, PDB accession number 1BL8), InsP3R1 (**g**, PDB accession number 3JAV), RyR1 (**h**, PDB accession number 3J8H), NavRh (**i**, PDB accession number 4DXW), CavAb (**j**, PDB accession number 4MVM), Cav1.1 domains I and III (**k**, PDB accession number 3JBR) and Cav1.1 domains II and IV (**l**, PDB accession number 3JBR). In each of the alignments, acidic residues located at or close to the selectivity filter region are shown as sticks for comparison. Notably, structures of Ca<sup>2+</sup>-permeable channels (**a**, **d**, **g**, **h**, **j–l**) display a high concentration of acidic residues in the outer pore region. In **a–c**, methionine residues close to the S6 bundle crossing are shown as sticks. Notably, the methionine at the lower gate points away from the pore

in TRPV1 (**a**), despite high sequence conservation in this region among TRPV channels (Extended Data Fig. 7). In Slo2.2 (**b**) and NavAb (**c**), methionine side chains occlude the lower gate as in TRPV6<sub>cryst</sub>, indicating that the closed conformation of the lower gate can be chemically similar for Na<sup>+</sup>-, K<sup>+</sup>- and Ca<sup>2+</sup>-selective channels. **m–o**, Comparison of calcium-binding sites in TRPV6<sub>cryst</sub> (**m**), the engineered voltage gated Ca<sup>2+</sup> channel CavAb (**n**) and the putative Ca<sup>2+</sup> site in Cav1.1 (**o**, domains I and III are shown). Residues constituting the selectivity filters are shown in stick representation. Ca<sup>2+</sup> ions are shown as green spheres. Sites 1 and 2 from TRPV6<sub>cryst</sub> overlap with the positions of sites 1 and 3 from CavAb, respectively. While it has been proposed that, owing to electrostatic repulsion, sites 1, 2 and 3 cannot be simultaneously occupied in CavAb, distances between Ca<sup>2+</sup>-binding sites in TRPV6<sub>cryst</sub> are sufficiently large such that they can be simultaneously occupied. The putative Ca<sup>2+</sup> site in Cav1.1 is near the equivalent location of site 2 in CavAb.



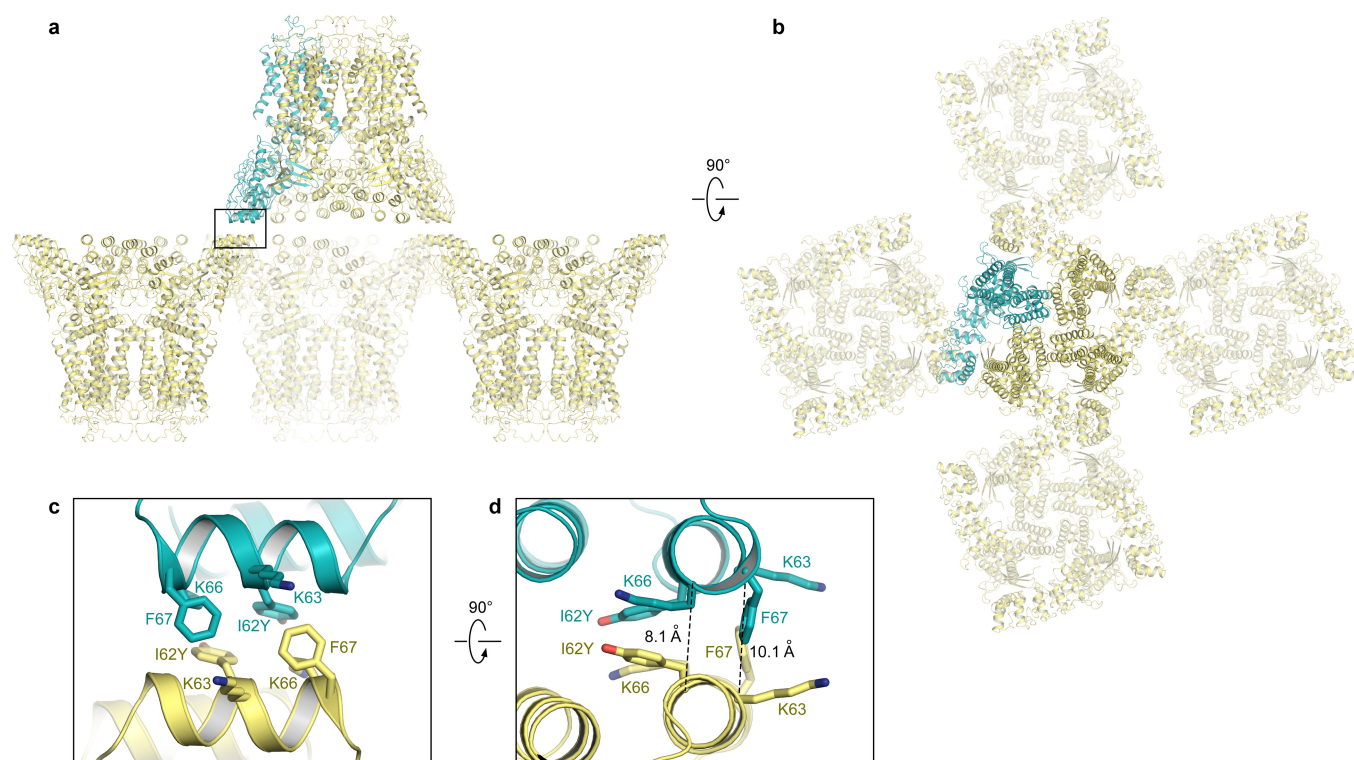
**Extended Data Figure 7 | Sequence alignment of rat TRPV subtypes.** Secondary structure elements are depicted above the sequence as cylinders ( $\alpha$ -helices), arrows ( $\beta$ -strands) and lines (loops). Dashed lines show residues in the TRPV6<sub>cryst</sub> construct not included in the TRPV6<sub>cryst</sub> structural model. Red boxes and a red arrow highlight substitution

mutations and the C-terminal truncation point in TRPV6<sub>cryst</sub> respectively (see Methods). The Y symbol marks the N-linked glycosylation site in the extracellular loop connecting S1 and S2 conserved in TRPV6 (and TRPV5) channels. The thick red line marks the location of the selectivity filter.



**Extended Data Figure 8 | Isothermal titration calorimetry analysis of TRPV6 interaction with  $\text{Gd}^{3+}$  and anomalous peak amplitudes.** **a**,  $\text{Gd}^{3+}$  in the syringe ( $700 \mu\text{M}$ ) was titrated into TRPV6 ( $6.38 \mu\text{M}$ ) loaded into the cell. Measurements were performed at  $25^\circ\text{C}$ . Top, the raw data for nineteen  $2\text{-}\mu\text{l}$  injections of  $\text{Gd}^{3+}$ . The area of each injection peak is equal to the total heat released from that injection. Bottom, the integrated heat per injection versus molar ratio. Binding of  $\text{Gd}^{3+}$  to TRPV6 was analysed using models with one and two types of binding site. A model with one type of binding site was not sufficient to explain the binding isotherm (blue line). In contrast, analyses of the binding isotherm using the model with two types of binding site, according to equation  $Q_i^{\text{tot}} = V_0 M_{\text{tot}} ((n_1 \Delta H_1 K_1 [X] / (1 + K_1 [X])) + (n_2 \Delta H_2 K_2 [X] / (1 + K_2 [X])))$ , where  $Q_i^{\text{tot}}$  is total heat after the  $i$ th injection,  $V_0$  is the volume of calorimetric cell,  $M_{\text{tot}}$  is the bulk concentration of protein,  $[X]$  is the free concentration of  $\text{Gd}^{3+}$ ,  $n_1$  and  $n_2$

are the numbers of type 1 and 2 sites,  $K_1$  and  $K_2$  are the observed equilibrium constants for each type of the sites and  $\Delta H_1$  and  $\Delta H_2$  are the corresponding enthalpy changes, satisfactorily described the data (red line), and the corresponding values of thermodynamic parameters are given in **b**. The values of  $\Delta G$  and  $T\Delta S$  were calculated using the following relationships:  $\Delta G = -RT \ln K$  and  $\Delta G = \Delta H - T\Delta S$ . **b**, Table showing the parameters of experimental data fitting to the model with two types of  $\text{Gd}^{3+}$ -binding site. The straightforward interpretation of the ITC results is that the ITC type 1 ( $n \approx 1$ ) and type 2 ( $n \approx 4$ ) sites represent the main (site 1) and recruitment sites identified crystallographically (Fig. 4e, f). Correspondingly, the affinity to  $\text{Gd}^{3+}$  for recruitment sites is  $\sim 10$  times lower than for site 1. **c**, Table showing anomalous peak amplitudes in  $\sigma$  calculated from data collected for  $\text{Ca}^{2+}$  ( $38\text{--}4.59 \text{ \AA}$ ),  $\text{Ba}^{2+}$  ( $38\text{--}4.59 \text{ \AA}$ ) and  $\text{Gd}^{3+}$  ( $38\text{--}4.59 \text{ \AA}$ ). No numbers are given if the peaks were not observed.



#### Extended Data Figure 9 | Crystal lattice contact of TRPV6<sub>cryst</sub>.

**a, b,** Two views of TRPV6<sub>cryst</sub> crystal packing in the *P42<sub>1</sub>2* space group. A single TRPV6<sub>cryst</sub> protomer in the asymmetric unit is shown in blue.

**c, d,** Close-up views of boxed region in **a**. Contacting residues are shown

in stick, and C<sub>α</sub>-C<sub>α</sub> distances are labelled in **d**. The crystal contact is apparently mediated by cation- $\pi$  and/or hydrogen bonding interactions between these residues. Crystals in the *P42<sub>1</sub>2* space group did not form when the native isoleucine was present at position 62.

Extended Data Table 1 | Data collection and refinement statistics

	Native	Ba <sup>2+</sup>	Ca <sup>2+</sup>	Gd <sup>3+</sup>	L630M-SeMet	SeMet
<b>Data Collection</b>						
Beamline	APS-24ID-C	APS-24ID-C	APS-24ID-C	APS-24ID-C	APS-24ID-E	APS-24ID-E
Space group	P4 <sub>2</sub> ,2	P4 <sub>2</sub> ,2	P4 <sub>2</sub> ,2	P4 <sub>2</sub> ,2	P4 <sub>2</sub> ,2	P4 <sub>2</sub> ,2
Cell dimensions	143.81	144.35	144.35	144.35	143.60	143.95
a, b, c, (Å)	143.81	144.35	144.35	144.35	143.60	143.95
	113.22	113.37	113.37	113.37	114.44	113.04
α, β, γ (°)	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90
Wavelength (Å)	0.9791	1.75	1.75	1.7101	0.9792	0.9792
Resolution (Å)*	44.48 - 3.25 (3.36 - 3.25)	49.56 - 3.85 (3.99 - 3.85)	49.56 - 3.65 (3.78 - 3.65)	50.00 - 3.80 (3.936 - 3.80)	40.00 - 7.20 (7.46 - 7.20)	40.00 - 5.00 (5.18 - 5.00)
Completeness (%)*	96.0 (94.7)	99.5 (95.4)	99.9 (99.8)	99.5 (97.3)	99.9 (100.0)	100.0 (100.0)
Redundancy*	8.7 (9.2)	15.4 (13.9)	26.5 (17.2)	11.4 (6.4)	16.3 (17.5)	13.7 (13.9)
σ *	16.9 (1.3)	15.2 (1.5)	25.0 (2.4)	19.4 (1.6)	27.3 (6.3)	21.8 (4.7)
R <sub>meas</sub> (%)*	9.8 (132.6)	13.1 (228.5)	10.6 (143.1)	8.9 (120.7)	20.7 (86.4)	19.2 (93.5)
CC <sub>1/2</sub>	99.8 (85.7)	98.0 (76.8)	99.5 (85.7)	99.9 (63.7)	98.5 (89.8)	98.3 (92.5)
<b>Refinement</b>						
Resolution (Å)*	44.48 - 3.25 (3.36 - 3.25)	49.56 - 3.85 (3.99 - 3.85)	50.00 - 3.65 (3.78 - 3.65)	49.56 - 3.80 (3.94 - 3.80)		
Completeness (%)	96 (93.8)	100 (99.9)	100 (99.9)	99 (96.9)		
Number of reflections	18531 (1724)	21705 (2187)	25439 (2521)	22443 (2170)		
R <sub>work</sub> /R <sub>free</sub>	0.273/0.289	0.291/0.326	0.276/0.281	0.298/0.321		
Number of atoms						
Total	4747	4775	4735	4759		
Ligand	16	19	18	17		
B-factor (Å <sup>2</sup> )						
Protein	120.5	143.8	135.1	144.23		
Ligand	77.27	136.75	24.14	178.86		
<b>RMS deviations</b>						
Bond length (Å)	0.003	0.002	0.003	0.002		
Bond angles (°)	0.7	0.62	0.62	0.63		
<b>Ramachandran</b>						
Favored (%)	93.6	92.9	93.7	92.8		
Allowed (%)	5.7	6.93	6.13	7.03		
Disallowed (%)	0.17	0.17	0.17	0.17		

\*Highest resolution shell in parentheses.

Five per cent of reflections were used for the calculation of R<sub>free</sub>.

# The first gravitational-wave source from the isolated evolution of two stars in the 40–100 solar mass range

Krzysztof Belczynski<sup>1</sup>, Daniel E. Holz<sup>2</sup>, Tomasz Bulik<sup>1</sup> & Richard O'Shaughnessy<sup>3</sup>

The merger of two massive (about 30 solar masses) black holes has been detected in gravitational waves<sup>1</sup>. This discovery validates recent predictions<sup>2–4</sup> that massive binary black holes would constitute the first detection. Previous calculations, however, have not sampled the relevant binary-black-hole progenitors—massive, low-metallicity binary stars—with sufficient accuracy nor included sufficiently realistic physics to enable robust predictions to better than several orders of magnitude<sup>5–10</sup>. Here we report high-precision numerical simulations of the formation of binary black holes via the evolution of isolated binary stars, providing a framework within which to interpret the first gravitational-wave source, GW150914, and to predict the properties of subsequent binary-black-hole gravitational-wave events. Our models imply that these events form in an environment in which the metallicity is less than ten per cent of solar metallicity, and involve stars with initial masses of 40–100 solar masses that interact through mass transfer and a common-envelope phase. These progenitor stars probably formed either about 2 billion years or, with a smaller probability, 11 billion years after the Big Bang. Most binary black holes form without supernova explosions, and their spins are nearly unchanged since birth, but do not have to be parallel. The classical field formation of binary black holes we propose, with low natal kicks (the velocity of the black hole at birth) and restricted common-envelope evolution, produces approximately 40 times more binary-black-holes mergers than do dynamical formation channels involving globular clusters<sup>11</sup>; our predicted detection rate of these mergers is comparable to that from homogeneous evolution channels<sup>12–15</sup>. Our calculations predict detections of about 1,000 black-hole mergers per year with total masses of 20–80 solar masses once second-generation ground-based gravitational-wave observatories reach full sensitivity.

We study the formation of coalescing black-hole binaries using the StarTrack population synthesis code<sup>16,17</sup>. This method has been updated to account for the formation of massive black-hole systems in isolated stellar environments. The new key factors include an observationally supported star-formation rate, chemical enrichment across cosmic time and a revised initial condition for evolution of binary stars. Hitherto, simulations have been unable to achieve the desired predictive power because of the limitations on the input physics (for example, limited metallicity range) and numerical accuracy. To ensure the dominant contribution from intrinsically rare low-metallicity star-forming environments are adequately sampled, we use a dense grid of metallicities (32 metallicities) with high precision (20 million binaries each).

Although binary population synthesis is dependent on a number of uncertain physical factors, there has been recent progress in reducing this uncertainty and understanding how it affects predictions. In light of this, we consider the following three models to encompass major sources of uncertainty (Methods): M1 represents our 'standard' classical formation model for double compact objects composed of two black holes (BH–BH), two neutron stars (NS–NS), or one of each (BH–NS);

M2 is our 'optimistic' model, in which Hertzsprung-gap stars may initiate and survive common-envelope evolution, leading to many more binaries being formed; and M3 is our 'pessimistic' model, in which black holes receive large natal kicks, which disrupts and thereby reduces the number of BH–BH progenitor binaries.

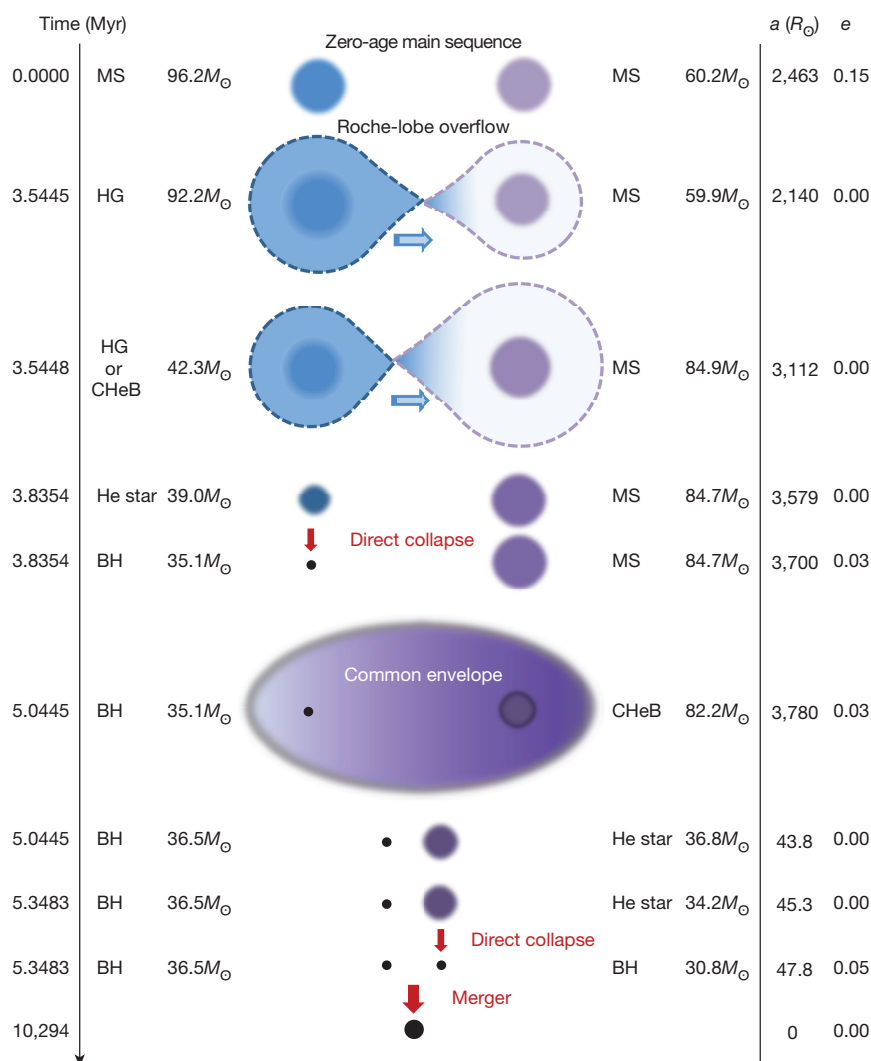
For each generated double compact object merger, with its intrinsic component masses and the redshift of the merger, we estimate the probability that such a merger would have been detectable in the first observing run (O1) of the Laser Interferometer Gravitational-Wave Observatory (LIGO) advanced detectors. We adopt a self-consistent model of evolution of stellar populations in the Universe<sup>3,4</sup>, and we take the representative noise curve for O1 (<https://dcc.ligo.org/LIGO-G1501223/public>) and assume 16 days of coincident science-quality observational time<sup>1</sup>.

In Fig. 1 we show the formation and evolution of a typical binary system that result in a merger with similar masses and at a similar time to GW150914. Stars that form such mergers are very massive ( $40M_{\odot}$ – $100M_{\odot}$ ;  $M_{\odot}$  is the mass of the Sun), and at the end of their lives they collapse directly to form black holes<sup>18</sup>. Because there is no associated supernova explosion, there is also no mass ejection. We allow 10% of the collapsing stellar mass to be emitted in neutrinos. If natal kicks are associated with asymmetric mass ejection (as in our standard model), then our prediction is that these massive black holes do not receive natal kicks and that their spin directions are the same as that of their progenitor collapsing stars. The binary evolution removes the hydrogen-rich envelope from both binary components, making both stars compact and luminous Wolf–Rayet stars before they collapse to black holes. The first binary interaction is a dynamically stable Roche-lobe overflow phase, whereas the second interaction consists of a common-envelope phase that produces a compact binary. After the common-envelope phase, the progenitor binary resembles two known high-mass X-ray binaries hosting massive black holes: IC10 X-1 and NGC 300 X-1 (ref. 19). A massive BH–BH binary (each with a mass of approximately  $30M_{\odot}$ ) is formed in approximately 5 Myr of evolution, with a relatively wide orbit (semi-major axis  $a \approx 50R_{\odot}$ ;  $R_{\odot}$  is the radius of the Sun), leading to a long time to coalescence of  $t_{\text{merger}} \approx 10$  Gyr. The accretion onto the first black hole in the common-envelope phase is only modest (approximately  $1.5M_{\odot}$ ), whereas accretion from stellar wind of its companion is rather small (less than  $0.1M_{\odot}$ ).

To investigate general aspects of the formation history of GW150914, we select a population of GW150914-like BH–BH mergers with a total redshifted mass of  $M_{\text{tot},z} = 54M_{\odot}$ – $73M_{\odot}$ , and then further restrict our sample to binaries that would be detectable in O1. The formation channels typical for these massive BH–BH mergers are summarized in Extended Data Table 1.

We find that the most likely progenitor of GW150914 consists of a primary star in the mass range  $40M_{\odot}$ – $100M_{\odot}$  and a secondary in the mass range  $40M_{\odot}$ – $80M_{\odot}$ . In our standard model, the binary formed

<sup>1</sup>Astronomical Observatory, Warsaw University, Ujazdowskie 4, 00-478 Warsaw, Poland. <sup>2</sup>Enrico Fermi Institute, Department of Physics, Department of Astronomy and Astrophysics, and Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA. <sup>3</sup>Center for Computational Relativity and Gravitation, Rochester Institute of Technology, Rochester, New York 14623, USA.



**Figure 1 | Example binary evolution leading to a BH-BH merger similar to GW150914.** A massive binary star ( $96M_{\odot}$  (blue) +  $60M_{\odot}$  (purple)) is formed in the distant past (2 billion years after Big Bang;  $z \approx 3.2$ ; top row), and after 5 million years of evolution forms a BH-BH system ( $37M_{\odot} + 31M_{\odot}$ ; second-last row). For the ensuing 10.3 billion years, this BH-BH system is subject to loss of angular momentum, with the orbital separation steadily decreasing, until the black holes coalesce at redshift  $z = 0.09$ . This example binary formed in a low-metallicity environment ( $Z = 0.03Z_{\odot}$ ). MS, main-sequence star; HG, Hertzsprung-gap star; ChEB, core-helium-burning star; BH, black hole;  $a$ , orbital semi-major axis;  $e$ , eccentricity.

in a low-metallicity environment ( $Z < 0.1Z_{\odot}$ ;  $Z_{\odot}$  is the metallicity of the Sun; see Extended Data Fig. 1) and either in the early Universe (2 Gyr after the Big Bang) or very recently (11 Gyr after the Big Bang).

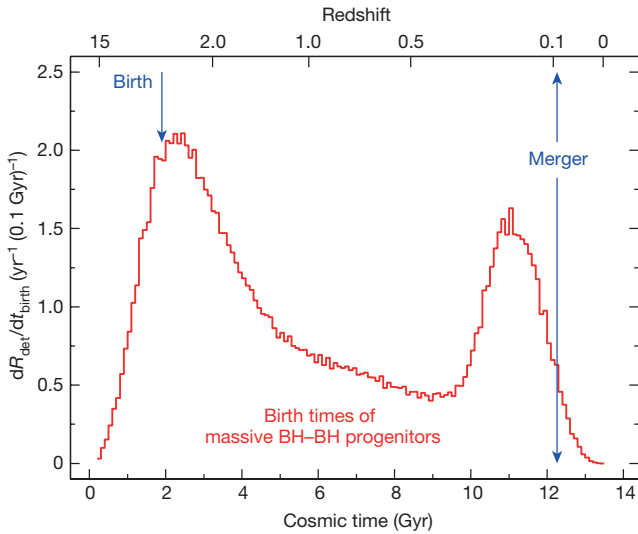
The distribution of birth times of these massive BH–BH mergers is bimodal (Fig. 2 and Extended Data Fig. 2), with a majority of systems originating from the distant past (55% of binaries; about 2 Gyr after the Big Bang, corresponding to  $z \approx 3$ ) and a smaller contribution from relatively young binaries (25%; formed about 11 Gyr after the Big Bang, corresponding to  $z \approx 0.2$ ). This bimodality arises from two naturally competing effects: on the one hand, most low-metallicity star formation occurs in the early Universe; on the other hand, in contrast to previous work<sup>3,4</sup>, significantly more low-metallicity star formation is currently expected to occur in the low-redshift Universe<sup>20</sup>. Therefore, as is the case with binary neutron stars, we anticipate a significant contribution to the present-day binary-black-hole merger rate from binary black holes formed in low-redshift, low-metallicity star-forming regions. The delay-time distribution of BH–BH binaries in our simulations follows a  $1/t$  distribution. The birth times therefore naturally pile up at low redshifts ( $z \approx 0.1\text{--}0.3$ ) and this gives rise to a low- $z$  peak (Extended Data Fig. 2a). However, the low-metallicity ( $Z < 0.1Z_{\odot}$ ) star formation responsible for the production of massive BH–BH mergers peaks at a redshift of  $z \approx 3$  (Extended Data Fig. 2b). The convolution of these two effects produces the bimodal birth-time distribution (Extended Data Fig. 2c).

These massive GW150914-like mergers consist of black holes with comparable masses. The vast majority (99.8%) of mergers are found with mass ratios in the range  $q = 0.7\text{--}1.0$  (Extended Data Fig. 3), with the mass ratio of GW150914 ( $q = 0.82^{+0.16}_{-0.21}$ ) falling near the centre of

the expected region. The formation of low-mass-ratio objects is suppressed because low-mass-ratio progenitors tend to merge during the first mass-transfer event when the more massive component overfills its Roche lobe<sup>21</sup>. However, with decreasing total merger mass, the mass ratio tends to lower values. In particular, for the lower mass bin of  $M_{\text{tot}} = 25M_{\odot} - 37M_{\odot}$ , mass ratios as low as  $q = 0.3$  are also found.

We now use our full sample of double compact object mergers to make predictions for the merger-rate density, detection rates and merger mass distribution. The results are shown in Fig. 3 and Extended Data Table 1, in which we compare them to the measured values inferred from O1 LIGO observations. We find an overall detection rate that is consistent with the detection of one significant candidate (GW150914) during the principal 16-day double coincident period (when both LIGO gravitational-wave interferometers are operating simultaneously) for our standard model (M1), but that is inconsistent for our other two models (optimistic M2 and pessimistic M3; more detail below).

The BH–BH merger rates inferred from the 16 days of O1 LIGO observations are in the range  $2\text{--}400\text{ Gpc}^{-3}\text{ yr}^{-1}$  (ref. 22). For comparison, we estimate the rate density of binary black holes from our population synthesis dataset. We consider the full population of binary black holes within a redshift of  $z=0.1$  (that is, not weighted by their detection probability) and calculate their average source-frame merger-rate density. We find a value of  $218\text{ Gpc}^{-3}\text{ yr}^{-1}$  for our standard model (M1), which is in good agreement with the inferred LIGO rate<sup>22</sup>. By contrast, our optimistic model (M2) predicts too many mergers, with a rate density of  $1,303\text{ Gpc}^{-3}\text{ yr}^{-1}$ , and our pessimistic model (M3)



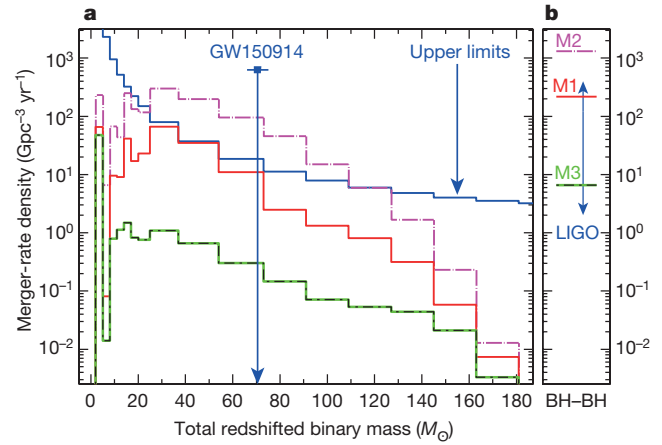
**Figure 2 | Birth times of GW150914-like progenitors across cosmic time.**  $dR_{\text{det}}/dt$  represents the contribution to the detection rate from binaries in a given 0.1-Gyr bin of birth time. Half of the binaries that form BH–BH mergers detectable in O1 with total redshifted mass in the range  $M_{\text{tot},z} = 54M_{\odot} - 73M_{\odot}$  were born within 4.7 Gyr of the Big Bang (corresponding to  $z > 1.2$ ). The birth and merger times of the binary depicted in Fig. 1 are marked in blue; this binary follows the most typical evolutionary channel for massive BH–BH mergers (BHBH1 in Extended Data Table 1). The merger redshift of GW150914 is  $z = 0.088$ . The bimodal shape of the distribution originates from a combination of the BH–BH delay-time distribution and the low-metallicity star-formation history (see Extended Data Fig. 2 for details).

is at the very bottom end of the allowable range with a predicted rate of  $6.6 \text{ Gpc}^{-3} \text{ yr}^{-1}$ . In our models, the BH–BH merger-rate density increases with redshift (Extended Data Fig. 4). This increase is modest; our predicted source-frame BH–BH merger-rate density would double if the cut-off redshift was increased from  $z = 0.1$  to  $z = 0.6$ .

The merger-rate density for the model with an optimistic common-envelope phase (M2) is an order of magnitude larger than the rate estimate from LIGO. This implies that unevolved massive stars (during main sequence and Hertzsprung gap) do not initiate/survive the common-envelope phase<sup>9,23</sup>. In our classical BH–BH formation scheme, only evolved stars (during core helium burning) with well-developed convective envelopes are allowed to initiate and survive the common-envelope phase.

Our predictions for the pessimistic model (M3) imply that large natal kicks (with average magnitudes of more than about  $400 \text{ km s}^{-1}$ ) are unlikely for massive black holes. This model predicts that an event such as GW150914 would happen only 1% of the time, with the detection of any BH–BH system happening less than 10% of the time (Table 1). In principle, this conclusion applies to the formation of only the first black hole in the binary, because large natal kicks lead to disruption of BH–BH progenitors while the binaries are wide. During the formation of the second black hole, the progenitor binaries are on very close orbits (Fig. 1) and are not disrupted by natal kicks. In Extended Data Fig. 4 we show a sequence of models with intermediate black-hole natal kicks; future observations may allow us to discriminate between these models and to constrain the natal-kick distribution. Future observations converging on M1 would indicate no natal kicks nor supernova explosions in massive black-hole formation<sup>18</sup>. A striking ramification of this is the prediction that hot and luminous Wolf–Rayet progenitors of massive black holes<sup>24</sup> should disappear from the sky as a result of direct collapse to a black hole (that is, with no supernova explosion). Targeted observational campaigns to search for such phenomena are already underway<sup>25</sup>.

Figure 3 shows the relative contribution to the overall merger-rate density associated with each bin of total redshifted merger mass  $M_{\text{tot},z}$ . For comparison, Fig. 3 also shows the fiducial sensitivity (see Methods) as



**Figure 3 | Comparison of merger rates and masses with O1 LIGO results.** Results are shown for standard (M1; red solid lines), optimistic common-envelope phase (M2; pink dash-dotted lines) and pessimistic large black-hole natal kicks (M3; green/black solid/dash-dotted line) models. **a**, Distribution of total redshifted binary mass. The merger-rate density of GW150914 ( $70.5M_{\odot}$ ) is indicated by the blue square (with 90% confidence interval in mass, and its vertical position arbitrary). The blue solid line shows the fiducial estimate of the sensitivity (or upper limits) of the 16-day O1 run. A comparison of the shapes of the blue and red lines suggests that the most likely detections for M1 are BH–BH mergers with masses in the range  $25M_{\odot} - 73M_{\odot}$ . NS–NS mergers (first bin) and BH–NS mergers (next five bins) are well below the estimated sensitivity and thus detections in O1 are not expected. The rate densities are in the detector rest frame. **b**, Comparison of the LIGO estimate of the BH–BH merger rate with our models. The LIGO value of  $2 - 400 \text{ Gpc}^{-3} \text{ yr}^{-1}$  (90% credible range) compares well with our standard (M1) and large black-hole natal kicks (M3) models. The rate densities are in the source rest frame. An updated version of Fig. 3, including additional gravitational-wave detections as they occur, can be found at <http://www.syntheticuniverse.org/stvsgwo.html>.

a function of mass, assuming equal-mass zero-spin binary black holes. Figure 3 demonstrates that the intersection of the strongly mass-dependent sensitivity and the intrinsic detectable mass distribution strongly favours sources with total redshifted masses of  $25M_{\odot} - 73M_{\odot}$ , consistent with recent work<sup>4</sup> and total redshifted mass of GW150914 ( $M_{\text{tot},z} = 70.5M_{\odot}$ ). In our simulations, the maximum intrinsic mass of a merging BH–BH binary is  $M_{\text{tot}} = 140M_{\odot}$ . When accounting for cosmological redshift

**Table 1 | Expected detection rate and number of detections**

Model	Merger type	O1 detection rate ( $\text{yr}^{-1}$ )	Number of detections in 16 days of O1
M1	All	63.18	2.770
	NS–NS	0.052	0.002
	BH–NS	0.231	0.010
	BH–BH	62.90	2.758
	GW150914	11.95	0.524
M2	All	476.1	20.87
	NS–NS	0.191	0.008
	BH–NS	0.796	0.035
	BH–BH	475.1	20.83
	GW150914	110.0	4.823
M3	All	1.985	0.087
	NS–NS	0.039	0.002
	BH–NS	0.014	0.001
	BH–BH	1.932	0.085
	GW150914	0.270	0.012

The first column indicates the model: standard (M1), optimistic common-envelope phase (M2), and large black-hole natal kicks (M3). The third column lists the expected detection rate  $R_{\text{det}}$  per unit double coincident time (both LIGO detectors operating at appropriate sensitivity), for a network comparable to O1, for different classes of mergers (indicated in the second column). The fourth column shows  $R_{\text{det}}T$ , where  $T = 16$  days is the analysis time relevant for the rate estimate for GW150914 (ref. 22). Entries for merger type ‘GW150914’ are for the subpopulation of BH–BH mergers with total redshifted mass in the range  $M_{\text{tot},z} = 54M_{\odot} - 73M_{\odot}$ .

( $M_{\text{tot},z} = (1+z)M_{\text{tot}}$ ) and taking into account the advanced O1 horizon redshift for this most massive binary ( $z=0.7$ ), the highest possible observed mass within O1 would be approximately  $240M_{\odot}$ .

Spin magnitudes and directions of merging black holes are potentially measurable by LIGO<sup>1</sup>. The second-born black hole in a BH–BH binary does not accrete mass, and its spin at merger is unchanged from its spin at birth. The first-born black hole, on the other hand, has a chance to accrete material from the stellar wind of the unevolved companion or during common-envelope evolution. However, because this is limited either by the very low efficiency of accretion from stellar winds or by inefficient accretion during common-envelope evolution<sup>26,27</sup>, the total accreted mass onto the first-born black hole is expected to be rather small (about  $1M_{\odot}$ – $2M_{\odot}$ ). This is insufficient to significantly increase the spin, and thus the spin magnitude of the first-born black hole at merger is within about 10% of its birth spin.

In our modelling, we assume that stars that are born in a binary have their spins aligned with the angular-momentum vector of the binary. If massive black holes do not receive natal kicks (for example, in our standard model M1), then our prediction is that black-hole spins are aligned during the final massive BH–BH merger. We note that our standard model includes natal kicks and mass loss for low-mass black holes (less than about  $10M_{\odot}$ ), and therefore BH–BH binaries with one or two low-mass black holes may show misalignment. Alternatively, binaries could be born with misalignment and retain it, misalignment could be caused by the third body or by interaction between the radiative envelope and the convective core<sup>28</sup>, or misalignment could result from a large natal kick on the second-born black hole. Several binaries are reported with misaligned spins<sup>29</sup>. Therefore, spin alignment of massive merging black holes suggests isolated field evolution, while misaligned spins do not elucidate formation processes.

As shown in Fig. 1, we find that the formation of massive BH–BH mergers is a natural consequence of isolated binary evolution. Our standard model (M1) of BH–BH mergers fully accounts for the observed merger-rate density and merger mass (Fig. 3), and for the mass ratio of two merging black holes (Extended Data Fig. 3) inferred from GW150914.

Our standard formation mechanism (M1) produces significantly more binary black holes than do alternative, dynamical channels associated with globular clusters. A recent study<sup>11</sup> suggests globular clusters could produce a typical merger rate of  $5 \text{ Gpc}^{-3} \text{ yr}^{-1}$ ; our standard model (M1) BH–BH merger-rate density is about 40 times larger:  $218 \text{ Gpc}^{-3} \text{ yr}^{-1}$ .

However, one non-classical isolated binary evolution channel involving rapidly rotating stars (homogeneous evolution) in very close binaries may also fully account for the formation of GW150914 (refs 12–15). In particular, typical rates of 1.8 detections in 16 days of O1 observations are found<sup>13</sup>, which is comparable to our prediction of 2.8 (Table 1). Only very massive BH–BH mergers with total intrinsic masses of more than about  $50M_{\odot}$  are formed in this model<sup>12,13</sup>, whereas our model predicts mergers with masses in a broader range, down to greater than about  $10M_{\odot}$ . Future LIGO observations of BH–BH mergers may allow us to discriminate between these two very different mass distributions/models.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 21 February; accepted 11 May 2016.**

- Abbott, B. P. *et al.* Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.* **116**, 061102 (2016).
- Belczynski, K. *et al.* The effect of metallicity on the detection prospects for gravitational waves. *Astrophys. J.* **715**, L138–L141 (2010).
- Dominik, M. *et al.* Double compact objects. III. Gravitational-wave detection rates. *Astrophys. J.* **806**, 263 (2015).
- Belczynski, K. *et al.* Compact binary merger rates: comparison with LIGO/Virgo upper limits. *Astrophys. J.* **819**, 108 (2016).
- Tutukov, A. V. & Yungelson, L. R. The merger rate of neutron star and black hole binaries. *Mon. Not. R. Astron. Soc.* **260**, 675–678 (1993).
- Lipunov, V. M., Postnov, K. A. & Prokhorov, M. E. Black holes and gravitational waves: possibilities for simultaneous detection using first-generation laser interferometers. *Astron. Lett.* **23**, 492–497 (1997).

- Nelemans, G., Yungelson, L. R. & Portegies Zwart, S. F. The gravitational wave signal from the Galactic disk population of binaries containing two compact objects. *Astron. Astrophys.* **375**, 890–898 (2001).
- Voss, R. & Tauris, T. M. Galactic distribution of merging neutron stars and black holes – prospects for short gamma-ray burst progenitors and LIGO/VIRGO. *Mon. Not. R. Astron. Soc.* **342**, 1169–1184 (2003).
- Belczynski, K., Taam, R. E., Kalogera, V., Rasio, F. A. & Bulik, T. On the rarity of double black hole binaries: consequences for gravitational wave detection. *Astrophys. J.* **662**, 504–511 (2007).
- Mennekens, N. & Vanbeveren, D. Massive double compact object mergers: gravitational wave sources and r-process element production sites. *Astron. Astrophys.* **564**, A134 (2014).
- Rodriguez, C. L., Chatterjee, S. & Rasio, F. A. Binary black hole mergers from globular clusters: masses, merger rates, and the impact of stellar evolution. *Phys. Rev. D* **93**, 084029 (2016).
- Marchant, P., Langer, N., Podsiadlowski, P., Tauris, T. M. & Moriya, T. J. A new route towards merging massive black holes. *Astron. Astrophys.* **588**, A50 (2016).
- de Mink, S. E. & Mandel, I. The chemically homogeneous evolutionary channel for binary black hole mergers: rates and properties of gravitational-wave events detectable by advanced LIGO. *Mon. Not. R. Astron. Soc.* <http://dx.doi.org/10.1093/mnras/stw1219> (2016).
- Eldridge, J. J. & Stanway, E. R. BPASS predictions for Binary Black-Hole Mergers. Preprint at <http://arxiv.org/abs/1602.03790> (2016).
- Woosley, S. E. The progenitor of GW150914. Preprint at <http://arXiv.org/abs/1603.00511> (2016).
- Belczynski, K., Kalogera, V. & Bulik, T. A comprehensive study of binary compact objects as gravitational wave sources: evolutionary channels, rates, and physical properties. *Astrophys. J.* **572**, 407–431 (2002).
- Belczynski, K. *et al.* Compact object merger with the StarTrack population synthesis code. *Astrophys. J. Suppl. Ser.* **174**, 223–260 (2008).
- Fryer, C. L. *et al.* Compact remnant mass function: dependence on the explosion mechanism and metallicity. *Astrophys. J.* **749**, 91 (2012).
- Bulik, T., Belczynski, K. & Prestwich, A. IC10 X-1/NGC300 X-1: the very immediate progenitors of BH–BH binaries. *Astrophys. J.* **730**, 140 (2011).
- Hirschauer, A. S. *et al.* ALFALFA discovery of the most metal-poor gas-rich galaxy known: AGC 198691. *Astrophys. J.* **822**, 108 (2016).
- Bulik, T., Gondek-Rosinska, D. & Belczynski, K. Expected masses of merging compact object binaries observed in gravitational waves. *Mon. Not. R. Astron. Soc.* **352**, 1372–1380 (2004).
- Abbott, B. P. *et al.* The rate of binary black hole mergers inferred from advanced LIGO observations surrounding GW150914. Preprint at <http://arxiv.org/abs/1602.03842> (2016).
- Pavlovskii, K. & Ivanova, N. Mass transfer from giant donors. *Mon. Not. R. Astron. Soc.* **449**, 4415–4427 (2015).
- Eldridge, J. J., Fraser, M., Smartt, S. J., Maund, J. R. & Crockett, R. M. The death of massive stars – II. Observational constraints on the progenitors of Type Ibc supernovae. *Mon. Not. R. Astron. Soc.* **436**, 774–795 (2013).
- Gerke, J. R., Kochanek, C. S. & Stanek, K. Z. The search for failed supernovae with the Large Binocular Telescope: first candidates. *Mon. Not. R. Astron. Soc.* **450**, 3289–3305 (2015).
- Ricker, P. M. & Taam, R. E. The interaction of stellar objects within a common envelope. *Astrophys. J.* **672**, L41–L44 (2008).
- MacLeod, M. & Ramirez-Ruiz, E. Asymmetric accretion flows within a common envelope. *Astrophys. J.* **803**, 41 (2015).
- Rogers, T. M., Lin, D. N. C., McElwaine, J. N. & Lau, H. H. B. Internal gravity waves in massive stars: angular momentum transport. *Astrophys. J.* **772**, 21 (2013).
- Albrecht, S. *et al.* The BANANA project. V. Misaligned and precessing stellar rotation axes in CV Velorum. *Astrophys. J.* **785**, 83 (2014).

**Acknowledgements** We are indebted to G. Wiktorowicz, W. Gladysz and K. Piszczek for their help with population synthesis calculations, and to H.-Y. Chen and Z. Doctor for their help with our LIGO/Virgo rate calculations. We thank the thousands of Universe@home users that have provided their personal computers for our simulations. We also thank the Hannover GW group for letting us use their ATLAS supercomputer. K.B. acknowledges support from the NCN grant Sonata Bis 2 (DEC-2012/07/E/ST9/01360). D.E.H. was supported by NSF CAREER grant PHY-1151836. D.E.H. also acknowledges support from the Kavli Institute for Cosmological Physics at the University of Chicago through NSF grant PHY-1125897 as well as an endowment from the Kavli Foundation. T.B. acknowledges support from the NCN grant Harmonia 6 (UMO-2014/14/M/ST9/00707). R.O'S. was supported by NSF grant PHY-1505629.

**Author Contributions** All authors contributed to the analysis and writing of the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.B. ([chrisbelczynski@gmail.com](mailto:chrisbelczynski@gmail.com)).

**Reviewer Information** Nature thanks M. Cantiello and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

Our Monte Carlo evolutionary modelling is performed with the StarTrack binary population synthesis code<sup>16</sup>. In particular, we incorporate a calibrated treatment of tidal interactions in close binaries<sup>17</sup>, a physical measure of the common envelope (CE) binding energy<sup>30,31</sup>, and a rapid-explosion supernova model that reproduces the observed mass gap between neutron stars and black holes (BHs)<sup>18,32</sup>. Our updated mass spectrum of BHs shows a strong dependence on the metallicity of the progenitor stars (Extended Data Fig. 5). In galaxies with metallicities similar to the Milky Way ( $Z = Z_{\odot} = 0.02$ ), BHs that formed out of single massive stars (initial mass  $M_{\text{ZAMS}} = 150 M_{\odot}$ ) reach a maximum mass of  $M_{\text{BH}} = 15 M_{\odot}$ , whereas, for very low metallicity ( $Z = 0.0001 = 0.005 Z_{\odot}$ ), the maximum mass becomes  $M_{\text{BH}} = 94 M_{\odot}$ . The above input physics represents our standard model (M1), which is representative of our classical formation scheme for double compact objects (BH–BH, BH–NS and NS–NS).

We have adopted specific values for a number of evolutionary parameters. Single stars are evolved with calibrated formulae based on detailed evolutionary calculations<sup>33</sup>. Massive star winds are adopted from detailed studies of radiation-driven mass loss<sup>34</sup>. For the Luminous Blue Variable phase, a high rate of mass loss is adopted ( $1.5 \times 10^{-4} M_{\odot} \text{ yr}^{-1}$ ). Binary interactions and, in particular, the stability of Roche-lobe overflow (RLOF) is judged on the basis of binary parameters: mass ratio, evolutionary stage of donor, response to mass loss, and behaviour of the orbital separation in response to mass transfer. The orbital separation is additionally affected by gravitational radiation, magnetic braking, and loss of angular momentum associated with systemic mass loss. During stable RLOF, we assume that half of the mass is accreted onto the companion, while the other half ( $1 - f_a = 0.5$ ) is lost with specific angular momentum ( $dJ/dt = j_{\text{loss}} J_{\text{orb}} / (M_{\text{don}} + M_{\text{acc}}) (1 - f_a) dM_{\text{RLOF}}/dt$  with scaling factor  $j_{\text{loss}} = 1.0$  where  $f_a$  is the fraction of the mass accreted,  $J_{\text{orb}}$  is the orbital angular momentum,  $M_{\text{don}}$  is the donor mass,  $M_{\text{acc}}$  is the accretor mass, and  $dM_{\text{RLOF}}/dt$  is the mass transfer rate; ref. 35). The CE is treated by considering the energy balance with fully effective conversion of orbital energy into envelope ejection (conversion efficiency  $\alpha = 1.0$ ), whereas the envelope binding energy for massive stars is calibrated by a parameter  $\lambda$ , which depends on star radius, mass and metallicity. For massive stars,  $\lambda \approx 0.1$  is adopted<sup>31</sup>. During CE evolution, compact objects accrete at 10% of the Bondi–Hoyle rate as estimated by recent hydrodynamical simulations<sup>26,27</sup>. Our CE evolution is instantaneous, so the time at the beginning and end of the CE phase is exactly the same (see Fig. 1); the time duration of the CE phase has no impact on our results.

We consider two extra variations of the input physics of binary evolution. In one model (M2), we test highly uncertain CE physics<sup>36</sup> and we allow for Hertzsprung-gap stars to initiate and survive CE evolution. This is an optimistic assumption, because these stars may not allow for CE evolution<sup>23</sup>, nor survive as a binary if a CE forms<sup>9</sup>. For comparison, in our standard model, we allow only evolved stars with a deep convective envelope (core-helium-burning stars) to survive a CE phase.

In the opposite extreme, we use a model (M3) in which BHs receive large natal kicks. In particular, each BH gets a natal kick with its components drawn from a Maxwellian distribution with a one-dimensional root-mean-square  $\sigma = 265 \text{ km s}^{-1}$ , independent of BH mass. Such large natal kicks are measured for Galactic pulsars<sup>37</sup>. This is a pessimistic assumption, because large natal kicks tend to disrupt BH–BH progenitor binaries. This assumption is not yet excluded on the basis of electromagnetic observations<sup>4</sup>. By contrast, in our standard model, BH natal kicks decrease with BH mass. In particular, for massive BHs that form through direct collapse of an entire star to a BH with no supernova explosion ( $M_{\text{BH}} \gtrsim 10 M_{\odot}$  for  $Z = Z_{\odot}$ ,  $M_{\text{BH}} \gtrsim 15 M_{\odot}$  for  $Z = 0.1 Z_{\odot}$  and  $M_{\text{BH}} \gtrsim 15 M_{\odot} - 30 M_{\odot}$  for  $Z = 0.01 Z_{\odot}$ ), we assume no natal kicks<sup>18</sup>. We also calculated a series of models with intermediate BH kicks (see Extended Data Fig. 4):  $\sigma = 200 \text{ km s}^{-1}$  (model M4),  $\sigma = 130 \text{ km s}^{-1}$  (model M5) and  $\sigma = 70 \text{ km s}^{-1}$  (model M6).

For each evolutionary model we compute  $2 \times 10^7$  massive binaries for each point on a grid of 32 sub-models covering a wide range of metallicities:  $Z = 0.0001, 0.0002, 0.0003, 0.0004, 0.0005, 0.0006, 0.0007, 0.0008, 0.0009, 0.001, 0.0015, 0.002, 0.0025, 0.003, 0.0035, 0.004, 0.0045, 0.005, 0.0055, 0.006, 0.0065, 0.007, 0.0075, 0.008, 0.0085, 0.009, 0.0095, 0.01, 0.015, 0.02, 0.025$  and  $0.03$ . We assume that stellar evolution at even lower metallicities proceeds in the same way as the evolution at  $Z = 0.005 Z_{\odot}$ . However, stars with very low metal content (for example, Population III) may evolve differently to metal-rich stars<sup>38</sup>.

Each sub-model is computed with initial distributions of orbital periods  $P$  (proportional to  $[\log(P)]^{-0.5}$ ), eccentricities  $e$  (proportional to  $e^{-0.42}$ ) and mass ratios  $q$  (proportional to  $q^0$ ) appropriate for massive stars<sup>39</sup>. We adopt an initial mass function that is close to flat for low-mass stars (proportional to  $M^{-1.3}$  for  $0.08 M_{\odot} \leq M < 0.5 M_{\odot}$  and to  $M^{-2.2}$  for  $0.5 M_{\odot} \leq M < 1.0 M_{\odot}$ ) and that is top-heavy for massive stars (proportional to  $M^{-2.3}$  for  $1.0 M_{\odot} \leq M \leq 150 M_{\odot}$ ), as guided by recent observations<sup>40</sup>. The adopted initial mass function generates higher BH–BH

merger-rate densities as compared with the steeper initial mass function (proportional to  $M^{-2.7}$  for  $1.0 M_{\odot} \leq M \leq 150 M_{\odot}$ ) adopted in previous studies<sup>4,41</sup>, because there are more BH–BH merger progenitors in our simulations<sup>42</sup>.

A moderate binary fraction ( $f_{\text{bi}} = 0.5$ ) is adopted for stars with masses  $M_{\text{ZAMS}} < 10 M_{\odot}$ , whereas we assume that all more massive stars are formed in binaries ( $f_{\text{bi}} = 1.0$ ), as indicated by recent empirical estimates<sup>39,43</sup>.

We adopt an extinction-corrected cosmic star-formation rate (SFR) based on numerous multi-wavelength observations<sup>44</sup>:

$$\text{SFR}(z) = 0.015 \frac{(1+z)^{2.7}}{1 + [(1+z)/2.9]^{5.6}} M_{\odot} \text{ Mpc}^{-3} \text{ yr}^{-1} \quad (1)$$

This SFR declines rapidly at high redshifts ( $z > 2$ ). This may be contrasted with some SFR models used previously<sup>45</sup>, which generated a greater number of stars at high redshifts. This revision will thus reduce the BH–BH merger-rate densities at all redshifts. Even though the formation of BH–BH binaries takes a very short time (about 5 Myr), the time to coalescence of two BHs may be very large (Fig. 1 and Extended Data Fig. 2).

In our treatment of chemical enrichment of the Universe, we follow the mean metallicity increase with cosmic time (since Big Bang until present). The mean metallicity as a function of redshift is:

$$\log[Z_{\text{mean}}(z)] = 0.5 + \log \left( \frac{y(1-R)}{\rho_b} \int_z^{20} \frac{97.8 \times 10^{10} \text{ SFR}(z')}{H_0 E(z')(1+z')} dz' \right)$$

with a return fraction  $R = 0.27$  (mass fraction of each generation of stars that is put back into the interstellar medium), a net metal yield  $y = 0.019$  (mass of new metal created and ejected into the interstellar medium by each generation of stars per unit mass locked in stars), a baryon density  $\rho_b = 2.77 \times 10^{11} \Omega_b h_0^2 M_{\odot} \text{ Mpc}^{-3}$  with  $\Omega_b = 0.045$  and  $h_0 = 0.7$ , a SFR given by Equation (1), and  $E(z) = \sqrt{\Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_{\Lambda}}$  with  $\Omega_{\Lambda} = 0.7$ ,  $\Omega_M = 0.3$ ,  $\Omega_k = 0$  and  $H_0 = 70.0 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The shape of the mean-metallicity dependence on redshift follows recent estimates<sup>44</sup>, although the level was increased by 0.5 dex to better fit observational data<sup>46</sup>. At each redshift, we assume a log-normal distribution of metallicity around the mean, with a standard deviation of  $\sigma = 0.5 \text{ dex}$  (ref. 47). Our prescription (Extended Data Fig. 6) produces more low-metallicity stars than previously<sup>41</sup>. Because BH–BH formation is enhanced at low-metallicity<sup>2</sup>, our new approach increases the predicted rate densities of BH–BH mergers.

Here we discuss caveats of evolutionary calculations. First, we consider only isolated binary evolution, and thus our approach is applicable to field stars in low-density environments. It is possible that dynamical interactions enhance BH–BH merger formation in dense globular clusters<sup>11</sup>, offering a completely independent channel.

Second, our predictions are based on a ‘classical’ theory of stellar and binary evolution for the modelling of massive stars that we have compiled, developed and calibrated over the last 15 years. We do not consider exotic channels for the formation of BH–BH mergers, such as the one from rapidly rotating stars in contact binaries<sup>48</sup>.

Third, our modelling includes only three evolutionary models: a standard model consisting of our best estimates for reasonable parameters (M1), as well as optimistic (M2) and pessimistic (M3) alternative models. The optimistic model consists of only one change from the standard model: we allow all stars beyond the main sequence to survive the CE phase. Alternatively, the pessimistic model also consists of only one change: larger BH natal kicks. We have not investigated other possible deviations from the standard model (for example, different assumptions of mass and angular-momentum loss during stable mass-transfer evolution) nor have we checked inter-parameter degeneracies (for example, models with large BH kicks and an optimistic CE phase). Precursor versions of these computationally demanding studies have already been performed<sup>49</sup>, albeit with low statistics and limited scope; these calculations indicate that our three models probably cover the range of interesting effects.

Fourth, our observations are severely statistically limited. We are attempting to draw inferences about our models on the basis of a single detection (GW150914).

It was argued<sup>50</sup> that the formation of GW150914 in isolated binary evolution requires a metallicity lower than  $0.5 Z_{\odot}$ . This argument was based on single stellar models<sup>51</sup>; stars in close binaries are subject to significant mass loss during RLOF/CE, and they form BHs with lower mass than BHs formed by single stars. Thus, in binaries, the metallicity threshold for massive BH formation is lower than in single stellar evolution. For example, formation of a single  $30 M_{\odot}$  BH requires  $Z < 0.25 Z_{\odot}$  (Extended Data Fig. 5, whereas formation of two such BHs in a binary requires  $Z < 0.10 Z_{\odot}$  (Extended Data Fig. 1). The value of this threshold depends on assumptions for the model of stellar evolution, winds and BH formation

processes. The physical models we have adopted yield a threshold of  $Z < 0.10Z_{\odot}$ , the same as that obtained with MESA (<http://mesa.sourceforge.net/>) for homogeneous stellar evolution<sup>12</sup>. Our model was calibrated using known masses of BHs and, in particular, we do not exceed  $15M_{\odot}$  for  $Z_{\odot}$  (the highest-mass stellar BH known in our Galaxy). By contrast, single stellar models used to derive the high metallicity threshold produce  $25M_{\odot}$  for  $Z_{\odot}$  (ref. 51). The highest threshold obtained with binary evolution was reported at the level of  $0.5Z_{\odot}$  (ref. 14). Such a high value of the metallicity threshold for the progenitor of GW150914 implies that stars at approximately solar metallicity ( $Z = 0.014$ ) produce BHs as massive as  $40M_{\odot}$  (ref. 14). This is neither supported nor excluded by available electromagnetic BH mass measurements (<https://stellarcollapse.org/bhmasses>).

In the following, we present calculation of the gravitational radiation signal. The output of StarTrack is a binary merger at a given time. We then calculate the gravitational waveform associated with this merger, and determine whether this binary would have been observable by LIGO in the O1 configuration<sup>3,4</sup>.

We model the full inspiral–merger–ringdown waveform of the binaries using the IMRPhenomD gravitational waveform template family<sup>52,53</sup>. This is a simple and fast waveform family that neglects the effects of spin (which are not relevant for GW150914). We consider a detection to be given by a threshold of  $\text{SNR} > 8$  in a single detector, and we use the fiducial O1 noise curve (<https://dcc.ligo.org/LIGO-G1501223/public>). We calculate the face-on, overhead SNR for each binary directly from equation (2) of ref. 3. We then calculate the luminosity distance at which this binary would be detected with  $\text{SNR} = 8$ . As the distance to the binary changes, the observer-frame (redshifted) mass also changes, and therefore calculating the horizon redshift requires an iterative process. Once this has been calculated, we then determine the predicted detection rates using equation (9) of ref. 3; the effects of the antenna power pattern are incorporated in the  $p_{\text{det}}$  term in this equation.

An estimate of fiducial advanced LIGO sensitivity during the 16-day GW150914 analysis is shown in Fig. 3. We estimate the sensitivity to coalescing compact binaries using a reference O1 noise curve. We assume that both detectors operate with the fiducial O1 noise curve, which is the same sensitivity we adopted to calculate compact binary detection rates. For comparison, this model agrees reasonably well with the ‘early-high’ sensitivity model<sup>54</sup>. Our expression is a 50th percentile upper limit, assuming no detections. The critical application of this expression is not related to its overall normalization; we are instead interested in its shape, which characterizes the strongly mass-dependent selection biases of LIGO searches.

Using these inputs, our fiducial estimate of the advanced LIGO sensitivity during the first 16 days of O1 for a specific mass bin  $\Delta M_i$  is:

$$R_{D,\Delta M_i,UL} = \frac{0.7}{V_{\Delta M_i} T}$$

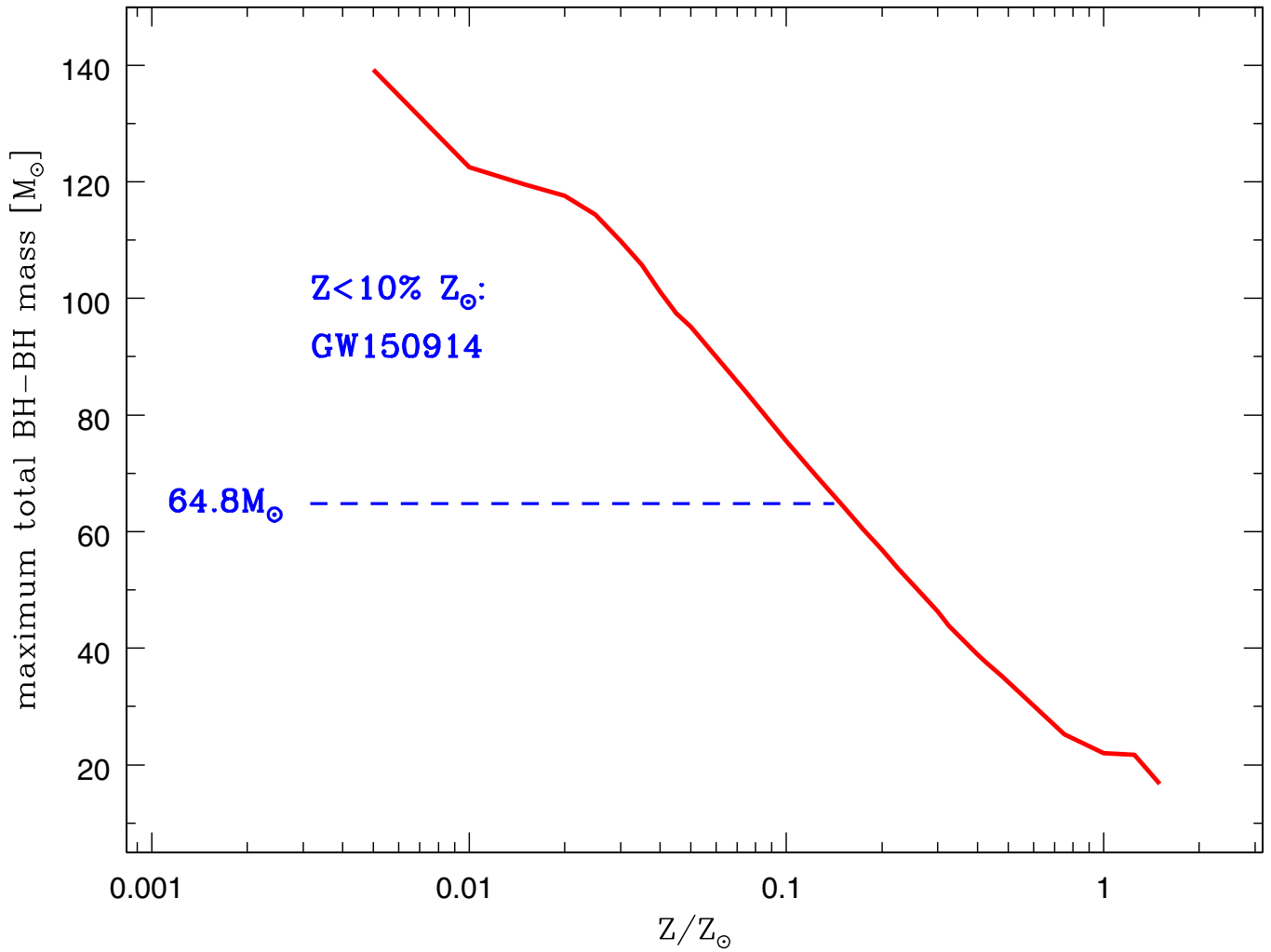
where  $T = 16$  days corresponds to the analysis of GW150914<sup>1</sup>, the volume:

$$V_{\Delta M_i} = \frac{1}{\Delta M_i} \int dM \int \frac{dz}{1+z} \frac{dV}{dz} p_{\text{det}}(w, M)$$

is the sensitive volume averaged over mass bin  $\Delta M_i$ , and  $p_{\text{det}}(w, M)$  is the orientation-averaged detection probability<sup>3,4</sup>. The function  $p_{\text{det}}(w, M)$  depends on the coalescing binary redshifted mass  $M$  through the maximum luminosity distance (‘horizon distance’) at which a source could produce a response of  $\text{SNR} > 8$  in a single detector through a projection parameter  $w$ , which is maximum ( $w = 1$ ) for a face-on, overhead source, and minimum ( $w = 0$ ) for sky locations and orientations where the LIGO detector has no response to the source. To calculate this distance, we adopt the IMRPhenomD gravitational waveforms<sup>52,53</sup> that we also used to estimate compact binary detection rates. Extended Data Fig. 7 shows our estimated horizon redshift as a function of the total redshifted binary merger mass for equal-mass mergers.

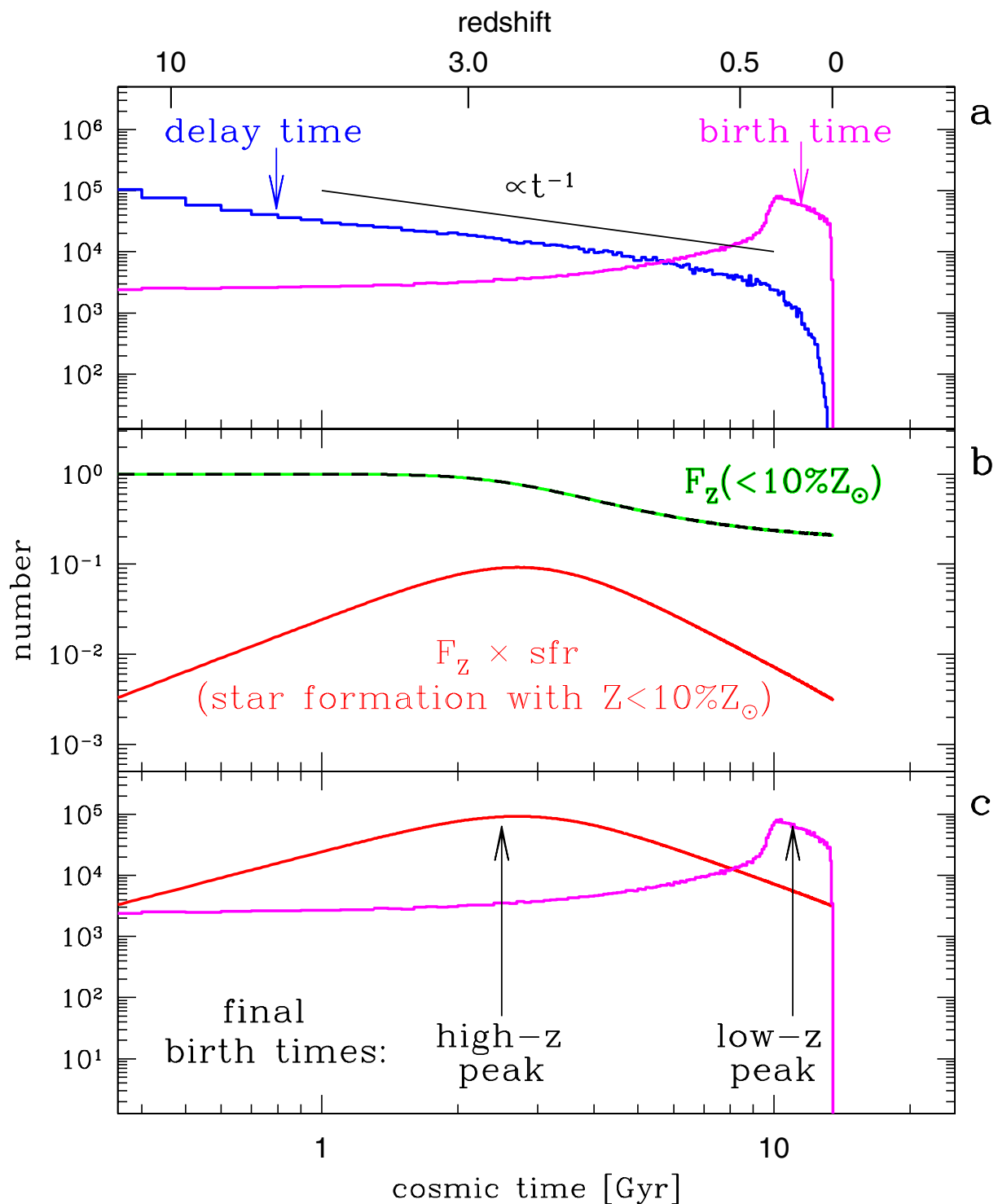
**Code availability.** We have opted not to release the population synthesis code StarTrack used to generate binary populations for this study.

30. Dominik, M. *et al.* Double compact objects. I. The significance of the common envelope on merger rates. *Astrophys. J.* **759**, 52 (2012).
31. Xu, X.-J. & Li, X.-D. Erratum: “On the binding energy parameter  $\lambda$  of common envelope evolution” (2010, *ApJ*, 716, 114). *Astrophys. J.* **722**, 1985–1988 (2010).
32. Belczynski, K., Wiktorowicz, G., Fryer, C. L., Holz, D. E. & Kalogera, V. Missing black holes unveil the supernova explosion mechanism. *Astrophys. J.* **757**, 91 (2012).
33. Hurley, J. R., Pols, O. R. & Tout, C. A. Comprehensive analytic formulae for stellar evolution as a function of mass and metallicity. *Mon. Not. R. Astron. Soc.* **315**, 543–569 (2000).
34. Vink, J. S. The theory of stellar winds. *Astrophys. Space Sci.* **336**, 163–167 (2011).
35. Podsiadlowski, P., Joss, P. C. & Hsu, J. J. L. Presupernova evolution in massive interacting binaries. *Astrophys. J.* **391**, 246–264 (1992).
36. Ivanova, N. *et al.* Common envelope evolution: where we stand and how we can move forward. *Astron. Astrophys. Rev.* **21**, 59 (2013).
37. Hobbs, G., Lorimer, D. R., Lyne, A. G. & Kramer, M. A statistical study of 233 pulsar proper motions. *Mon. Not. R. Astron. Soc.* **360**, 974–992 (2005).
38. Szécsi, D. *et al.* Low-metallicity massive single stars with rotation. Evolutionary models applicable to I Zwicky 18. *Astron. Astrophys.* **581**, A15 (2015).
39. Sana, H. *et al.* Binary interaction dominates the evolution of massive stars. *Science* **337**, 444–446 (2012).
40. Bastian, N., Covey, K. R. & Meyer, M. R. A universal stellar initial mass function? A critical look at variations. *Annu. Rev. Astron. Astrophys.* **48**, 339–389 (2010).
41. Dominik, M. *et al.* Double compact objects. II. Cosmological merger rates. *Astrophys. J.* **779**, 72 (2013).
42. de Mink, S. E. & Belczynski, K. Merger rates of double neutron stars and stellar origin black holes: the impact of initial conditions on binary evolution predictions. *Astrophys. J.* **814**, 58 (2015).
43. Duchêne, G. & Kraus, A. Stellar multiplicity. *Annu. Rev. Astron. Astrophys.* **51**, 269–310 (2013).
44. Madau, P. & Dickinson, M. Cosmic star-formation history. *Annu. Rev. Astron. Astrophys.* **52**, 415–486 (2014).
45. Strolger, L.-G. *et al.* The Hubble Higher z Supernova Search: supernovae to  $z \approx 1.6$  and constraints on Type Ia progenitor models. *Astrophys. J.* **613**, 200–223 (2004).
46. Vangioni, E. *et al.* The impact of star formation and gamma-ray burst rates at high redshift on cosmic chemical evolution and reionization. *Mon. Not. R. Astron. Soc.* **447**, 2575–2587 (2015).
47. Dvorkin, I., Silk, J., Vangioni, E., Petitjean, P. & Olive, K. A. The origin of dispersion in DLA metallicities. *Mon. Not. R. Astron. Soc.* **452**, L36–L40 (2015).
48. Almeida, L. A. *et al.* Discovery of the massive overcontact binary VFTS 352: evidence for enhanced internal mixing. *Astrophys. J.* **812**, 102 (2015).
49. O’Shaughnessy, R., Kalogera, V. & Belczynski, K. Mapping population synthesis event rates on model parameters. II. Convergence and accuracy of multidimensional fits. *Astrophys. J.* **667**, 1048–1058 (2007).
50. Abbott, B. P. *et al.* Astrophysical implications of the binary black hole merger GW150914. *Astrophys. J.* **818**, L22 (2016).
51. Spera, M., Mapelli, M. & Bressan, A. The mass spectrum of compact remnants from the PARSEC stellar evolution tracks. *Mon. Not. R. Astron. Soc.* **451**, 4086–4103 (2015).
52. Khan, S. *et al.* Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D* **93**, 044007 (2016).
53. Husa, S. *et al.* Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *Phys. Rev. D* **93**, 044006 (2016).
54. Abbott, B. P. *et al.* Prospects for observing and localizing gravitational-wave transients with Advanced LIGO and Advanced Virgo. *Living Rev. Relativ.* **19**, 1 (2013).
55. Villante, F. L., Serenelli, A. M., Delahaye, F. & Pinsonneault, M. H. The chemical composition of the sun from helioseismic and solar neutrino data. *Astrophys. J.* **787**, 13 (2014).



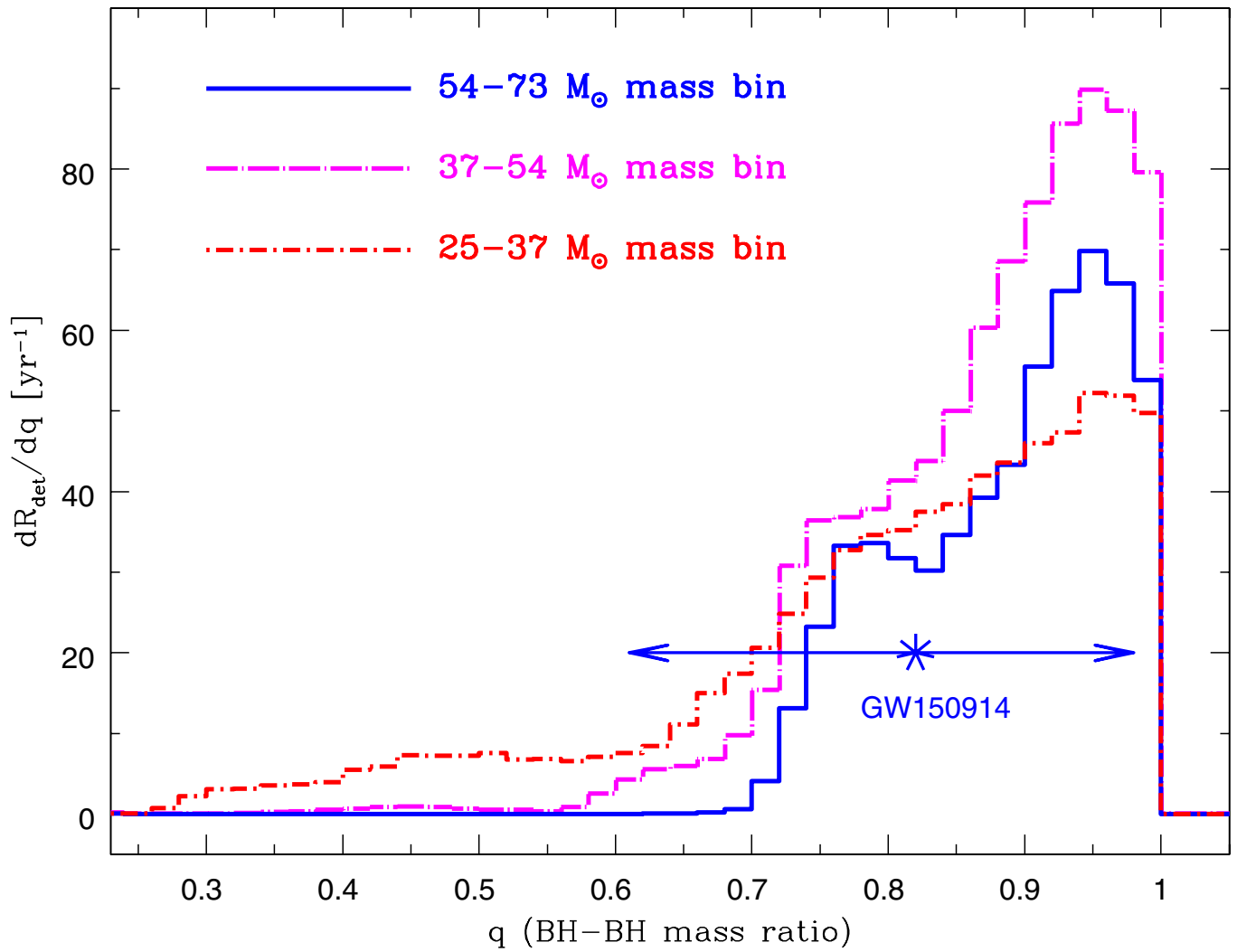
**Extended Data Figure 1 | Maximum total mass of BH–BH mergers as a function of metallicity.** Binary stars at metallicities  $Z < 0.1Z_{\odot}$  can form BH–BH mergers that are more massive than  $M_{\text{tot}} = 64.8M_{\odot}$ . This suggests that GW150914 was formed in a low-metallicity environment, assuming it is a product of classical isolated binary evolution. The total binary-

maximum BH–BH mass is not a simple sum of maximum BH masses resulting from single stellar evolution; this is a result of mass loss during the RLOF and CE evolution phases in the formation of massive BH–BH mergers (Fig. 1).



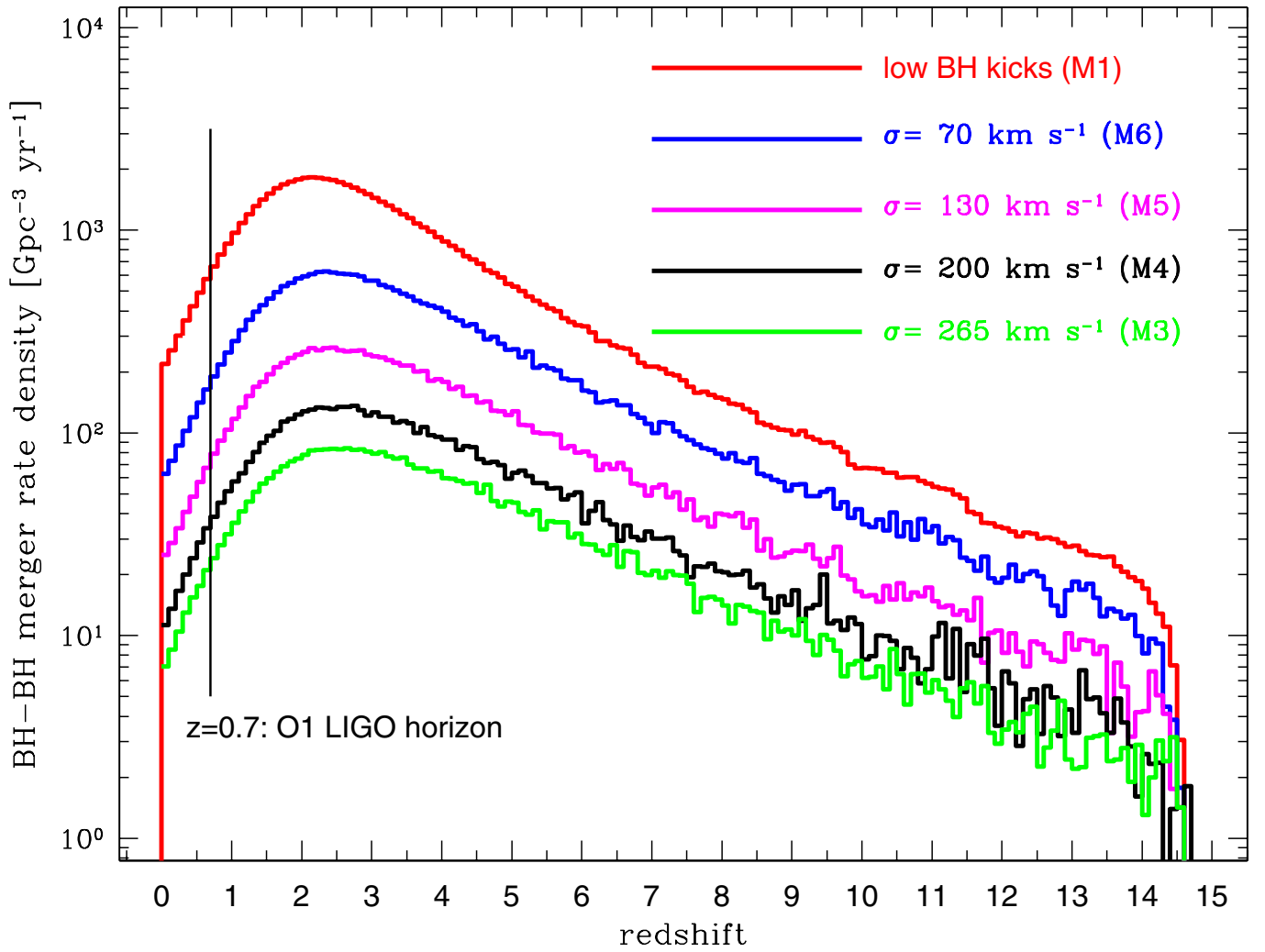
**Extended Data Figure 2 | Emergence of a bimodal birth-time distribution.** **a**, BH binaries follow an intrinsic power-law delay-time distribution (proportional to  $t^{-1}$ ). The birth time ( $t_{\text{birth}} = t_{\text{merger}} - t_{\text{delay}}$ ) is inverted compared to the delay-time distribution (blue line), with the spread caused by allowing the merger time ( $t_{\text{merger}}$ ) to fall anywhere within the O1 LIGO horizon:  $z = 0-0.7$ ; this generates a peak corresponding to BH-BH progenitors born late with short delay times. **b**, Massive BH-BH

binaries are formed by only low-metallicity stars ( $Z < 0.10Z_\odot$ ). The fraction of all stars that form at such low  $Z$  ( $F_Z$ ) decreases with cosmic time, making low- $Z$  star formation (in units of  $M_\odot \text{ Mpc}^{-3} \text{ yr}^{-1}$ ) peak at early cosmic time. sfr, star-formation rate. **c**, The final birth-time distribution for massive BH-BH mergers is a convolution of the intrinsic birth times and the low-metallicity star-formation rate.



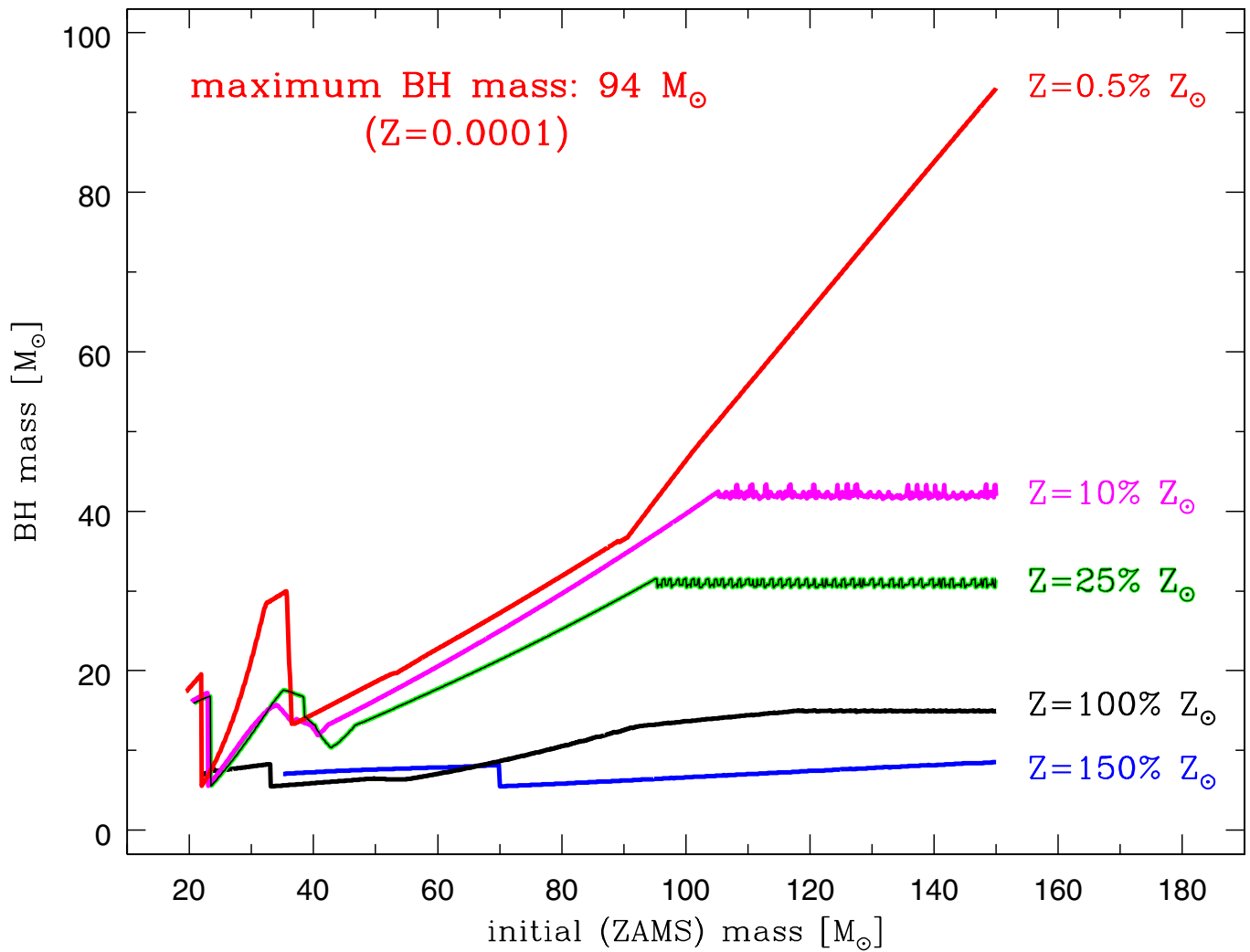
**Extended Data Figure 3 | Predicted distribution of BH-BH merger mass ratios.**  $dR_{\text{det}}/dq$  is the contribution to the detection rate,  $R_{\text{det}}$ , from binaries within a given 0.02 bin in mass ratio,  $q$ . Standard model (M1) detector-frame mass ratio is shown. BH-BH binaries prefer mass ratios of

$q \gtrsim 0.7$ , with a prominent peak near comparable-mass systems. GW150914, with  $q = 0.82^{+0.16}_{-0.21}$  (90% credible range) and a total redshifted mass of  $M_{\text{tot},z} = 70.5 M_{\odot}$ , falls within the expected region.



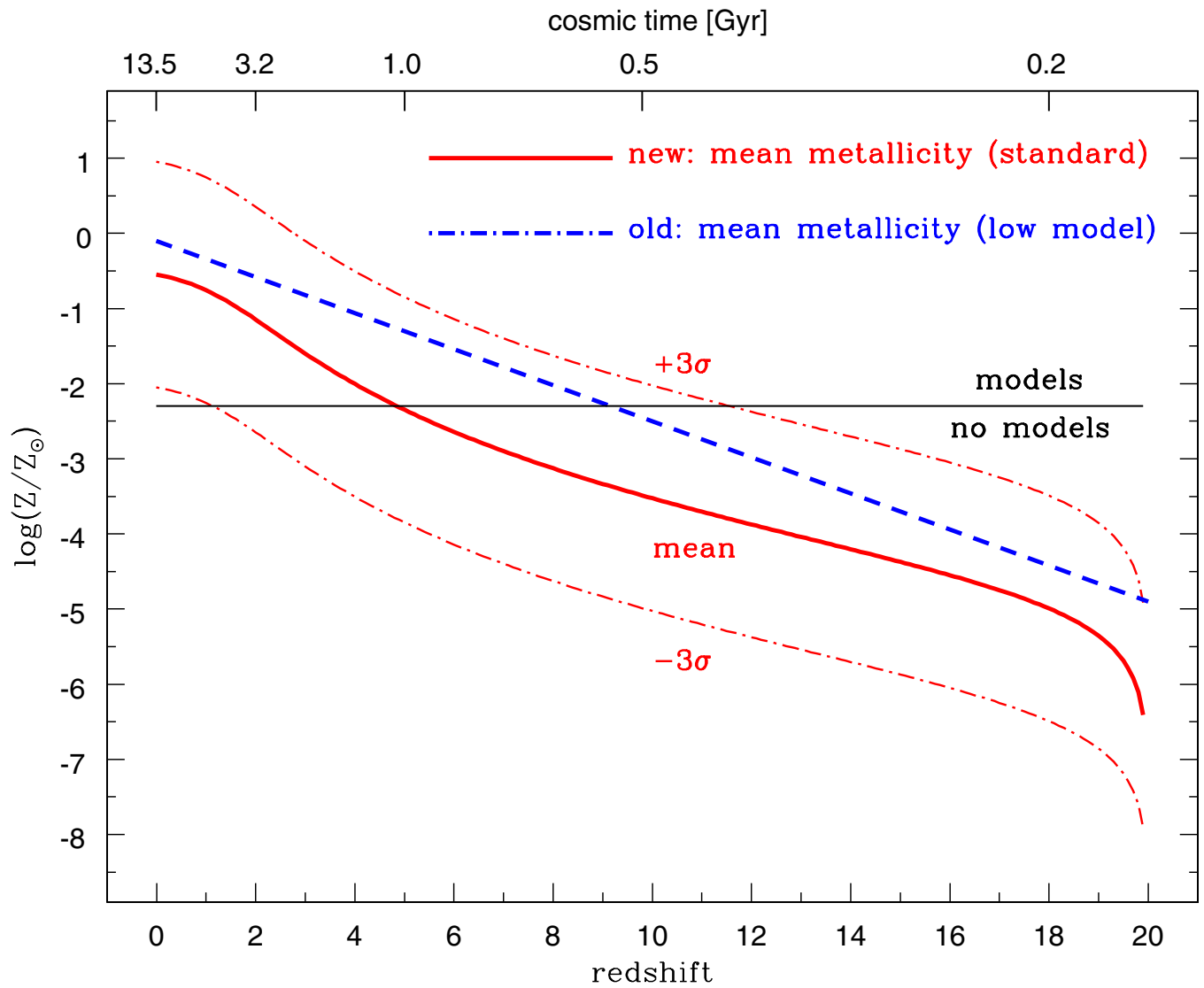
**Extended Data Figure 4 | Source-frame merger-rate density for BH-BH binaries as a function of redshift.** The red line shows the results from our standard model (M1); in this model, massive BHs do not get natal kicks. A sequence of models with increasing BH natal kicks (models M6, M5, M4, M3) is shown. The rate density decreases with increasing natal kick strength described by a Maxwellian distribution with a one-dimensional

root mean square deviation of  $\sigma$ . The local merger-rate density ( $z < 0.1$ ) changes from  $218 \text{ Gpc}^{-3} \text{ yr}^{-1}$  (M1) to  $63 \text{ Gpc}^{-3} \text{ yr}^{-1}$  (M6),  $25 \text{ Gpc}^{-3} \text{ yr}^{-1}$  (M5),  $11 \text{ Gpc}^{-3} \text{ yr}^{-1}$  (M4) and  $6.6 \text{ Gpc}^{-3} \text{ yr}^{-1}$  (M3). The LIGO estimate ( $2\text{--}400 \text{ Gpc}^{-3} \text{ yr}^{-1}$ ) encompasses all of these models. We mark the O1 LIGO detection horizon ( $z = 0.7$ ; see Extended Data Fig. 7).



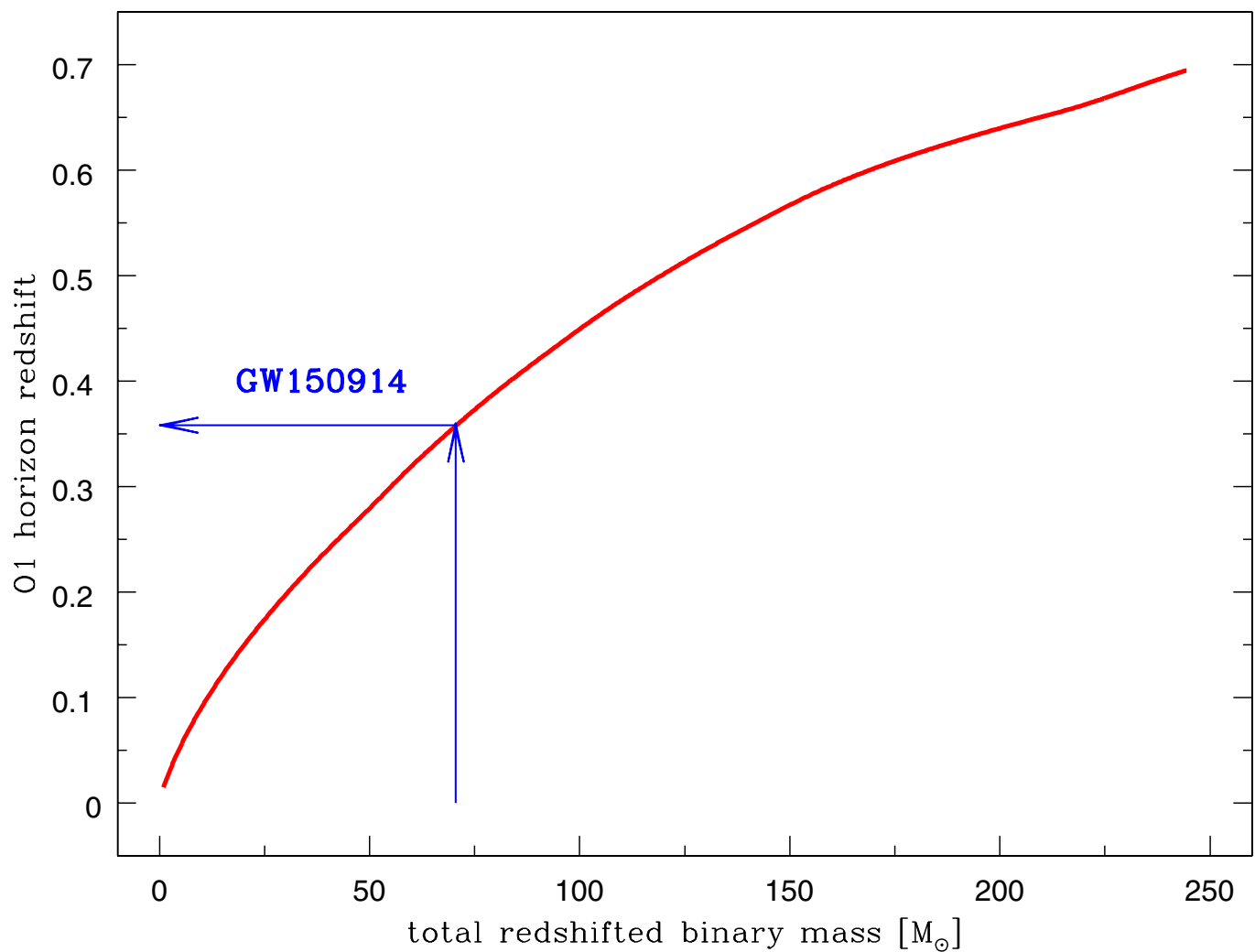
**Extended Data Figure 5 | BH mass as a function of initial star mass, for a range of metallicities.** These results show calculations for single star evolution with no binary interactions. Our updated models of BH formation show a general increase of BH mass with initial progenitor star mass. There is strong dependence of BH mass on the

chemical composition of the progenitor. For example, the maximum BH mass increases from  $10M_{\odot}$ – $15M_{\odot}$  for high-metallicity progenitors ( $Z = 1.5Z_{\odot}$ – $1Z_{\odot}$ ) to  $94M_{\odot}$  for low-metallicity progenitors ( $Z = 0.005Z_{\odot}$ ). The formation of a single  $30M_{\odot}$  BH requires a metallicity of  $Z \leq 0.25Z_{\odot}$ . ZAMS, zero-age main sequence.



**Extended Data Figure 6 | Mean-metallicity evolution of the Universe with redshift.** It is assumed that at each redshift the metallicity distribution is log-normal with a standard deviation of  $\sigma = 0.5$  dex. The blue line denotes the mean-metallicity evolution adopted in previous studies. The new relation generates more low-metallicity stars at all

redshifts. We mark the line above which we can make predictions ( $\log(Z/Z_{\odot}) = -2.3$ ,  $Z_{\odot} = 0.02$ ; ref. 55) based on actual evolutionary stellar models adopted in our calculations. Below this line we assume that stars produce BH–BH mergers in the same way as in the case of our lowest available model.



**Extended Data Figure 7 | Horizon redshift for the first advanced LIGO observational run (O1).** Horizon is given as a function of the total redshifted binary merger mass (assuming equal-mass mergers). For the highest-mass mergers found in our simulations ( $M_{\text{tot},z} = 240M_{\odot}$ ), the horizon redshift is  $z_{\text{hor}} = 0.7$ . For GW150914 ( $M_{\text{tot},z} = 70.5M_{\odot}$ ), the horizon redshift is  $z_{\text{hor}} = 0.36$ .

Extended Data Table 1 | Formation channels of massive BH–BH mergers (M1)

Channel		Evolutionary sequence			all [%]	high- <i>z</i>	mid- <i>z</i>	low- <i>z</i>
BHBH1	MT1(2-1)	BH1	CE2(14-4;14-7)	BH2	79.481	38.045	18.673	22.763
BHBH2	MT1(4-1)	BH1	CE2(14-4;14-7)	BH2	13.461	10.766	1.101	1.594
BHBH3	MT1(4-4)	CE2(4/7-4;7-7)	BH1	BH2	5.363	4.852	0.194	0.317
Other		additional combinations			1.696	0.625	0.421	0.649

The first two columns identify evolutionary sequences leading to the formation of BH–BH mergers with  $M_{\text{tot},z} = 54M_{\odot} - 73M_{\odot}$ . The third column lists the formation efficiency. The last three columns list the formation efficiency of BH–BH progenitors born at  $z > 1.12$ ,  $1.12 < z < 0.34$ ,  $z < 0.34$ . Notation: stable mass transfer (MT), common envelope (CE), BH formation (BH) initiated by either the primary star (1) or the secondary star (2). In parentheses we give the evolutionary stage of stars during MT/(pre-;post-)CE: main sequence (1), Hertzsprung gap (2), core helium-burning (4), helium star (7) or BH (14), with the primary star listed first.

# Real-time dynamics of lattice gauge theories with a few-qubit quantum computer

Esteban A. Martinez<sup>1\*</sup>, Christine A. Muschik<sup>2,3\*</sup>, Philipp Schindler<sup>1</sup>, Daniel Nigg<sup>1</sup>, Alexander Erhard<sup>1</sup>, Markus Heyl<sup>2,4</sup>, Philipp Hauke<sup>2,3</sup>, Marcello Dalmonte<sup>2,3</sup>, Thomas Monz<sup>1</sup>, Peter Zoller<sup>2,3</sup> & Rainer Blatt<sup>1,2</sup>

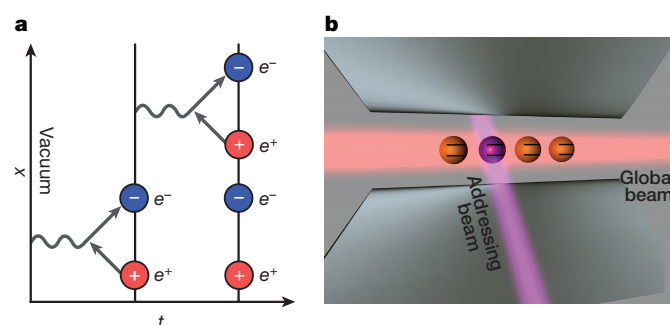
Gauge theories are fundamental to our understanding of interactions between the elementary constituents of matter as mediated by gauge bosons<sup>1,2</sup>. However, computing the real-time dynamics in gauge theories is a notorious challenge for classical computational methods. This has recently stimulated theoretical effort, using Feynman's idea of a quantum simulator<sup>3,4</sup>, to devise schemes for simulating such theories on engineered quantum-mechanical devices, with the difficulty that gauge invariance and the associated local conservation laws (Gauss laws) need to be implemented<sup>5–7</sup>. Here we report the experimental demonstration of a digital quantum simulation of a lattice gauge theory, by realizing (1 + 1)-dimensional quantum electrodynamics (the Schwinger model<sup>8,9</sup>) on a few-qubit trapped-ion quantum computer. We are interested in the real-time evolution of the Schwinger mechanism<sup>10,11</sup>, describing the instability of the bare vacuum due to quantum fluctuations, which manifests itself in the spontaneous creation of electron–positron pairs. To make efficient use of our quantum resources, we map the original problem to a spin model by eliminating the gauge fields<sup>12</sup> in favour of exotic long-range interactions, which can be directly and efficiently implemented on an ion trap architecture<sup>13</sup>. We explore the Schwinger mechanism of particle–antiparticle generation by monitoring the mass production and the vacuum persistence amplitude. Moreover, we track the real-time evolution of entanglement in the system, which illustrates how particle creation and entanglement generation are directly related. Our work represents a first step towards quantum simulation of high-energy theories using atomic physics experiments—the long-term intention is to extend this approach to real-time quantum simulations of non-Abelian lattice gauge theories.

Small-scale quantum computers exist today in the laboratory as programmable quantum devices<sup>14</sup>. In particular, trapped-ion quantum computers<sup>13</sup> provide a platform allowing a few hundred coherent quantum gates to act on a few qubits, with a clear roadmap towards scaling up these devices<sup>4,15</sup>. This provides the tools for universal digital quantum simulation<sup>16</sup>, where the time evolution of a quantum system is approximated as a stroboscopic sequence of quantum gates<sup>17</sup>. Here we show how this technology can be used to simulate the real-time dynamics of a minimal model of a lattice gauge theory, realizing the Schwinger model<sup>8,9</sup> as a one-dimensional quantum field theory with a chain of trapped ions (Fig. 1).

Our few-qubit demonstration is a first step towards simulating real-time dynamics in gauge theories: such simulations are fundamental for the understanding of many physical phenomena, including thermalization after heavy-ion collisions and pair creation studied at high-intensity laser facilities<sup>6,18</sup>. Although existing classical numerical methods such as quantum Monte Carlo have been remarkably successful for describing equilibrium phenomena, no systematic techniques exist to tackle the dynamical long-time behaviour of all but very small

systems. In contrast, quantum simulations aim at the long-term goal of solving the specific yet fundamental class of problems that currently cannot be tackled by these classical techniques. The digital approach we employ here is based on the Hamiltonian formulation of gauge theories<sup>9</sup>, and enables direct access to the system wavefunction. As we show below, this allows us to investigate entanglement generation during particle–antiparticle production, emphasizing a novel perspective on the dynamics of the Schwinger mechanism<sup>2</sup>.

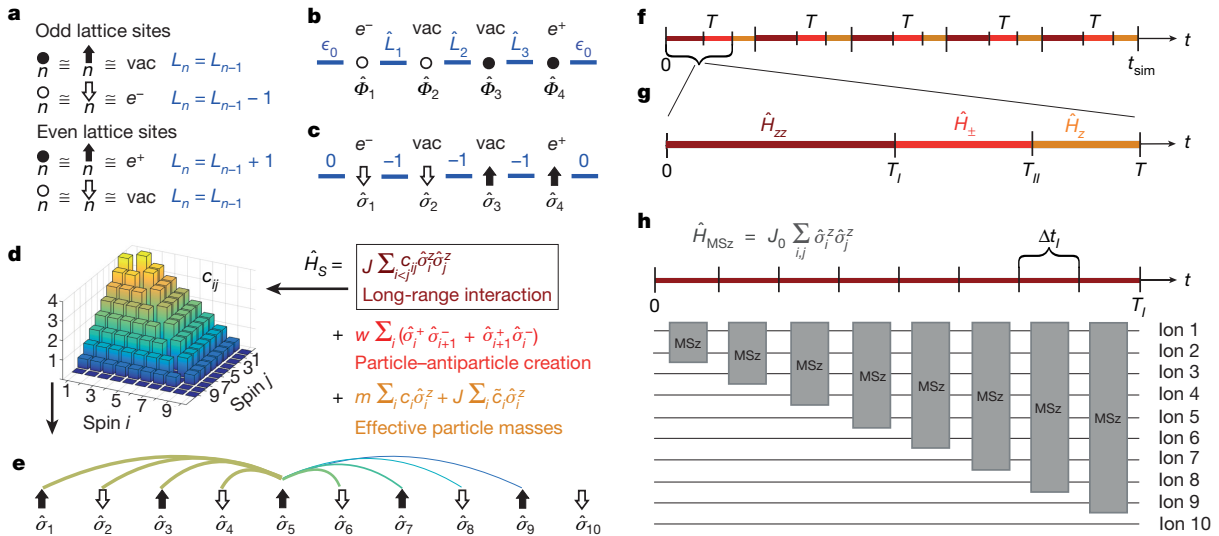
Digital quantum simulations described in the present work are conceptually different from, and fundamentally more challenging than, previously reported condensed-matter-motivated simulations of spin and Hubbard-type models<sup>4,19,20</sup>. In gauge theories, local symmetries lead to the introduction of dynamical gauge fields obeying a Gauss law<sup>6</sup>. Formally, this crucial feature is described by local symmetry generators  $\{\hat{G}_i\}$  that commute with the Hamiltonian of the system  $[\hat{H}, \hat{G}_i] = 0$  and restrict the dynamics to a subspace of physical states  $|\Psi_{\text{physical}}\rangle$  which satisfy  $\hat{G}_i|\Psi_{\text{physical}}\rangle = q_i|\Psi_{\text{physical}}\rangle$ , where  $q_i$  are background charges. We will be interested in the case  $q_i = 0$  for all  $i$  (see Methods). Realizing such a constrained dynamics on a quantum simulator is demanding and has been the focus of theoretical research<sup>6,7,11,21–24</sup>. Instead, to optimally use the finite resources represented by a few qubits of existing quantum hardware, we encode the gauge degrees of freedom in a long-range interaction between the fermions (electrons and positrons), which can be implemented efficiently on our experimental platform. This allows us to explore quantum simulation of coherent real-time



**Figure 1 | Quantum simulation of the Schwinger mechanism.** **a**, The instability of the vacuum due to quantum fluctuations is one of the most fundamental effects in gauge theories. We simulate the coherent real-time dynamics of particle–antiparticle creation by realizing the Schwinger model (one-dimensional quantum electrodynamics) on a lattice, as described in the main text. **b**, The experimental setup for the simulation consists of a linear Paul trap, where a string of  $^{40}\text{Ca}^+$  ions is confined. The electronic states of each ion, depicted as horizontal lines, encode a spin  $|\uparrow\rangle$  or  $|\downarrow\rangle$ . These states can be manipulated using laser beams (see Methods for details).

<sup>1</sup>Institute for Experimental Physics, University of Innsbruck, 6020 Innsbruck, Austria. <sup>2</sup>Institute for Quantum Optics and Quantum Information of the Austrian Academy of Sciences, 6020 Innsbruck, Austria. <sup>3</sup>Institute for Theoretical Physics, University of Innsbruck, 6020 Innsbruck, Austria. <sup>4</sup>Physics Department, Technische Universität München, 85747 Garching, Germany.

\*These authors contributed equally to this work.



**Figure 2 | Encoding Wilson's lattice gauge theories in digital quantum simulators.** Matter fields, represented by one-component fermion fields  $\hat{\phi}_n$  at sites  $n$ , interact via equation (1) with gauge variables defined on the links connecting the sites. **a**, Unoccupied odd (occupied even) sites, represented by filled (empty) circles, indicate the presence of an electron (positron). **b**, Gauge variables (shown as horizontal blue thick lines) are represented by operators  $\hat{L}_n$  with integer eigenvalues  $L_n = 0, \pm 1, \dots, \pm \infty$ . **c**, By mapping the fields  $\hat{\phi}_n$  to Pauli operators  $\hat{\sigma}_n$ , we obtain a spin model (the spins are represented by filled/empty arrows). In this language, the Gauss law governing the interaction of fermions and gauge variables reads  $\hat{L}_n - \hat{L}_{n-1} = \frac{1}{2}[\hat{\sigma}_n^z + (-1)^n]$ , where  $\hat{\sigma}^z$  is the diagonal Pauli matrix. The realization of the Schwinger model on a small-scale device requires an optimized use of resources. We achieve this by eliminating the gauge fields at the cost of obtaining a model with long-range couplings (and additional

dynamics with four qubits, exemplified here by the creation of electron-positron pairs (Fig. 1).

To this end, we experimentally study the Schwinger model, which describes quantum electrodynamics in one dimension. This model is extensively used as a testbed for lattice gauge theories as it shares many important features with quantum chromodynamics, including confinement, chiral symmetry breaking, and a topological theta vacuum<sup>6</sup>. In the Kogut-Susskind Hamiltonian formulation of the Schwinger model<sup>8,9</sup>,

$$\hat{H}_{\text{lat}} = -iw \sum_{n=1}^{N-1} [\hat{\phi}_n^\dagger e^{i\theta_n} \hat{\phi}_{n+1} - \text{h.c.}] + J \sum_{n=1}^{N-1} \hat{L}_n^2 + m \sum_{n=1}^N (-1)^n \hat{\phi}_n^\dagger \hat{\phi}_n \quad (1)$$

describes the interaction of fermionic field operators  $\hat{\phi}_n$  at sites  $n = 1 \dots N$  with gauge fields that are represented by the canonically commuting operators  $[\theta_n, \hat{L}_m] = i\delta_{n,m}$ .  $\hat{L}_n$  and  $\hat{\theta}_n$  correspond to the electromagnetic field and vector potential on the connection between sites  $n$  and  $n+1$ . The latter can be eliminated by a gauge transformation (see Methods). The fields  $\hat{\phi}_n$  represent Kogut-Susskind fermions (Fig. 2), where the presence of an electron (positron) is mapped to an unoccupied odd (occupied even) lattice site, allowing for a convenient incorporation of particles and antiparticles in a single fermion field. Accordingly, the third term in equation (1), representing the rest mass  $m$ , obtains a staggered sign. The first term corresponds to the creation and annihilation of particle-antiparticle pairs, and the second term reflects the energy stored in the electric field. Their energy scales  $w = 1/(2a)$  and  $J = g^2 a/2$  depend on the lattice spacing  $a$  and the fermion light coupling constant  $g$ . We use natural units  $\hbar = c = 1$ ;

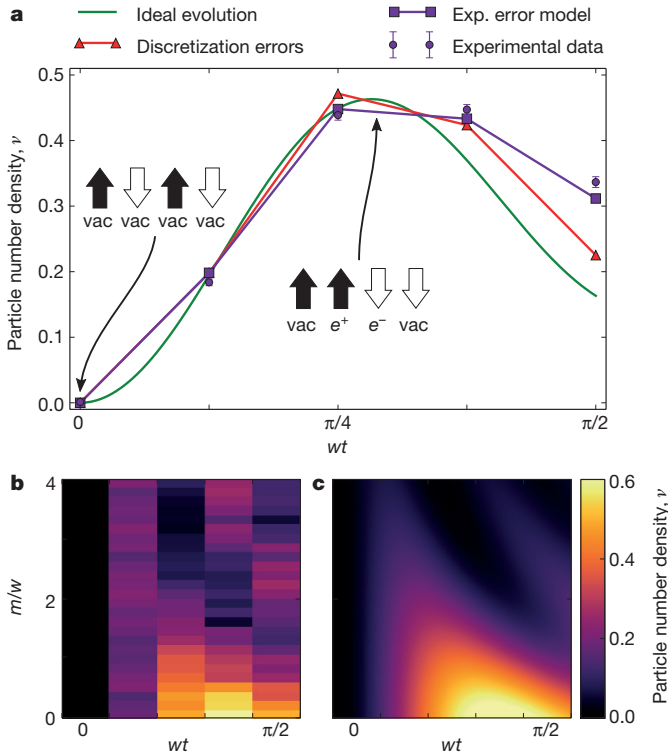
local terms). More specifically, the Gauss law determines the gauge fields for a given matter configuration and background field  $\epsilon_0$ . The elimination of the operators  $\hat{L}_n$  transforms the original model with nearest-neighbour terms into a pure spin model with long-range couplings that corresponds to the Coulomb interaction between the charged particles. **d**, Coupling matrix of the resulting interactions for  $N = 10$ , along with the total spin Hamiltonian  $\hat{H}_S$ . For illustration, **e** shows the couplings involving the fifth spin. The colours (and thicknesses) of lines represent the different interaction strengths  $c_{ij}$  according to the matrix shown in **d**. For implementing  $\hat{H}_S$  in a scalable and efficient way, we introduce time steps of length  $T$  (**f**), each subdivided into three sections (**g**). In each of these (length not to scale), one of the three parts of  $\hat{H}_S$  is realized as explained in Methods. **h**, The protocol for realizing  $\hat{H}_{zz}$  for  $N = 10$ . The ions interact according to the Mølmer-Sørensen (MS) Hamiltonian  $\hat{H}_{\text{MSz}}$ . During each short time window of length  $\Delta t_i$ , a different set of ions is coupled by  $\hat{H}_{\text{MSz}}$ .

therefore,  $a$  and  $t$  have the dimension of length, while  $w, J, m$  and  $g$  have the dimension of inverse length.

To realize the model using trapped ions, we map the fermionic operators  $\hat{\phi}_n$  to spin operators (Fig. 2a) by a Jordan-Wigner transformation<sup>12</sup>, which converts the short-range hopping in equation (1) into nearest-neighbour spin flip terms. In this formulation, the Gauss law takes the form  $\hat{L}_n - \hat{L}_{n-1} = \frac{1}{2}[\hat{\sigma}_n^z + (-1)^n]$ , where  $\sigma_n$  are the Pauli matrices. This law is the lattice version of the continuum law  $\nabla E = \rho$ , where  $\rho$  is the charge density. As illustrated in Fig. 2c, the Gauss law completely determines the electric fields for a given spin configuration and choice of background field. Following ref. 12, we use this constraint to eliminate the operators  $\hat{L}_n$  from the dynamics, adapting a scheme that has previously proven advantageous for numerical calculations<sup>25</sup> to a quantum simulation experiment, where the Gauss law is fulfilled by construction.

The elimination of the gauge fields maps the original problem to a spin model with long-range interactions that reflect the Coulomb interactions between the simulated particles. This allows an efficient use of resources, since  $N$  spins can be used to simulate  $N$  particles and their accompanying  $N-1$  gauge fields. However, as shown in Fig. 2d, the required couplings and local terms have a very unusual distance and position dependence. The challenge has thus been moved from engineering a constrained dynamics of  $2N-1$  quantum systems on a gauge-invariant Hilbert space to the realization of an exotic and asymmetric interaction of  $N$  spins.

Our platform is ideally suited for this task, since long-range interactions and precise single qubit operations are available in trapped-ion systems. These capabilities allow us to realize the required interactions by means of a digital quantum simulation scheme<sup>17</sup>. To this end, the desired Hamiltonian,  $H = \sum_{k=1}^K H_k$ , is split into  $K$  parts that can be directly implemented and are applied separately in subsequent time



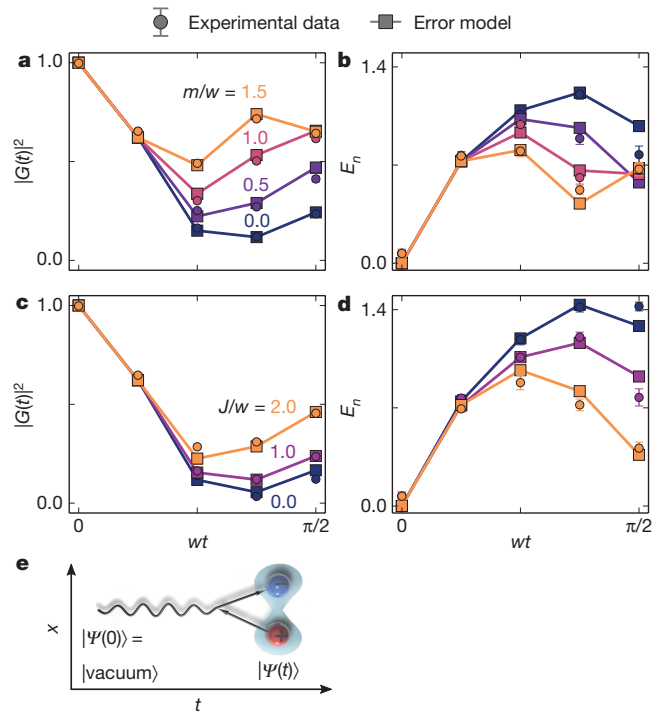
**Figure 3 | Time evolution of the particle number density,  $\nu$ .** **a**, We show the ideal evolution under the Schwinger Hamiltonian  $\hat{H}_S$  shown in Fig. 2d, the ideal evolution considering time discretization errors (see Fig. 2), the expected evolution including an experimental (exp.) error model (see Methods) and the experimental data for electric field energy  $J = w$  and particle mass  $m = 0.5w$  (see equation (1)). After postselection of the experimental data (see Methods), the remaining populations are  $\{86 \pm 2, 79 \pm 1, 73 \pm 1, 69 \pm 1\}\%$  after  $\{1, 2, 3, 4\}$  time steps (averaged over all data sets). Error bars correspond to standard deviations estimated from a Monte Carlo bootstrapping procedure. The insets show the initial state of the simulation (left inset), corresponding to the bare vacuum with particle number density  $\nu = 0$ , as well as one example of a state containing one pair (right inset), that is, a state with  $\nu = 0.5$ , represented as filled/empty arrows as in Fig. 2. **b**, Experimental data and **c**, theoretical prediction for the evolution of the particle number density  $\nu$  as a function of the dimensionless time  $wt$  and the dimensionless particle mass  $m/w$ , with  $J = w$ .

windows. By repeating the sequence multiple times, the resulting time evolution of the system  $U(t)$  closely resembles an evolution where the individual parts of the Hamiltonian act simultaneously, as can be shown using the Suzuki–Lie–Trotter expansion:

$$U(t) = e^{-i\hat{H}t} = \lim_{n \rightarrow \infty} \left( \bigotimes_{k=1}^K e^{-i\hat{H}_k t/n} \right)^n$$

Our scheme is depicted in Fig. 2f–h. It allows for an efficient realization of the required dynamics and implements the coupling matrix shown in Fig. 2d, e with a minimal number of time steps, scaling only linearly in the number of sites  $N$ . The scheme is therefore scalable to larger systems. A discussion of finite size effects can be found in Methods.

We realize the simulation in a quantum information processor based on a string of  $^{40}\text{Ca}^+$  ions confined in a macroscopic linear Paul trap (Fig. 1b). There, each qubit is encoded in the electronic states  $|\downarrow\rangle = 4S_{1/2}$  (with magnetic quantum number  $m = -1/2$ ),  $|\uparrow\rangle = 3D_{5/2}$  ( $m = -1/2$ ) of a single ion. The energy difference between these states is in the optical domain, so the state of the qubit can be manipulated using laser light pulses. More specifically, a universal set of high-fidelity quantum operations is available, consisting of collective rotations around the equator of the Bloch sphere, addressed rotations around the  $z$  axis and



**Figure 4 | Time evolution of the vacuum persistence amplitude and entanglement.** We show the square of the vacuum persistence amplitude  $|G(t)|^2$  (the Loschmidt echo), which quantifies the decay of the unstable vacuum, and the logarithmic negativity  $E_n$ , a measure of the entanglement between the left and the right halves of the system. **a**, **b**, The time evolution of  $|G(t)|^2$  (**a**) and  $E_n$  (**b**) for different values of the particle mass  $m$  and fixed electric field energy  $J = w$ , where  $w$  is the rate of particle–antiparticle creation and annihilation (compare equation (1)), as a function of the dimensionless time  $wt$ . **c**, **d**, The time evolution of  $|G(t)|^2$  (**c**) and  $E_n$  (**d**) changes for different values of  $J$  and fixed particle mass  $m = 0$ . Circles correspond to the experimental data and squares connected by solid lines to the expected evolution assuming an experimental error model explained in Methods. Error bars correspond to standard deviations estimated from a Monte Carlo bootstrapping procedure. **e**, Illustration of the creation of a particle–antiparticle pair starting from the bare vacuum state.

entangling Mølmer–Sørensen (MS) gates<sup>26</sup>. With a sequence of these gates, arbitrary unitary operations can be implemented<sup>27</sup>. Thus, we are able to simulate any Hamiltonian evolution, and in particular the interactions required here, by means of digital quantum simulation techniques, as shown in Fig. 2. Each of the implemented time evolutions consists of a sequence of over 200 quantum gates (see Extended Data Fig. 3). In order to realize the non-local interactions  $H_{zz}$  and  $H_{\pm}$  with their specific long-range interactions, we use global MS entangling gates together with a spectroscopic decoupling method to tailor the range of the interaction. For the decoupling, the population of the ions that are not involved in the specific operations are shelved into additional electronic states that are not affected by the light for the entangling operations (see Methods). The local terms in  $H_z$  correspond to  $z$  rotations that are directly available in our set of operations. The strength of all terms can be tuned by changing the duration of the laser pulses corresponding to the physical operations.

Within our scheme, a wide range of fundamental properties in one-dimensional lattice gauge theories can be studied. To demonstrate our approach, we concentrate on simulating the coherent quantum real-time dynamics of the Schwinger mechanism, that is, the creation of particle–antiparticle pairs out of the bare vacuum  $|\text{vacuum}\rangle$ , where matter is entirely absent (see Methods). After initializing the system in this state, which corresponds to the ground state for  $m \rightarrow \infty$  (Fig. 3a), we apply  $\hat{H}_S$  (Fig. 2d) for different masses and coupling strengths. As a first step, we measure the particle number density

$\nu(t) = \frac{1}{2N} \sum_{i=1}^N \langle (-1)^i \hat{\sigma}_i^z(t) + 1 \rangle$  generated after a simulated time evolution of duration  $t$ . The value  $\nu = 0.5$  corresponds to a state containing on average one pair (Fig. 2b). As Fig. 3c shows, an initial phase of rapid pair creation is followed by a reduction of  $\nu(t)$  due to recombination effects. The measured evolution shows excellent agreement with theoretical predictions, assuming uncorrelated dephasing with an error probability  $p = 0.038$  per qubit and per step, as explained in Methods. In Fig. 3b, we probe the particle–antiparticle generation for a broad range of masses  $m$ . Larger values of  $m$  increase the energy cost of pair production and thus lead to faster oscillations with a suppressed magnitude (see also Methods and Extended Data).

Our platform allows direct measurements of the vacuum persistence amplitude and of the generated entanglement. The vacuum persistence amplitude  $G(t) = \langle \text{vacuum} | e^{-iHt} | \text{vacuum} \rangle$  quantifies the decay of the unstable vacuum (see Methods). The associated probability  $|G(t)|^2$  shown in Fig. 4a, c, also known as the Loschmidt echo, is important in contexts such as quantum chaos<sup>28</sup> and dynamical critical phenomena far from equilibrium<sup>29</sup>.

The vacuum decay continuously produces entanglement, as particles and antiparticles are constantly generated and propagate away from each other, thus correlating distant parts of the system. Entanglement plays a crucial role in the characterization of dynamical processes in quantum many-body systems, and its analysis permits us to quantify the quantum character of the generated correlations. To this end, we reconstruct the density matrix after each time step by full state tomography, and evaluate the entanglement of one half of the system with the other by calculating the logarithmic negativity. This quantity is an entanglement measure for mixed states<sup>30</sup>, which is defined as the sum of the negative eigenvalues of the partially transposed density matrix. The entanglement between two contiguous blocks of our spin system is equivalent to the entanglement in the simulated fermionic system described by equation (1), that is, including the gauge fields (C.A.M. *et al.*, manuscript in preparation). In Fig. 4b, d, we show the real-time dynamics of the logarithmic negativity for different parameter regimes. Entanglement between the two halves of the system is due to the presence of a pair distributed across them. Accordingly, less entanglement is produced for increasing particle masses  $m$  and field energies  $J$ . The latter has a stronger influence, as it not only raises the energy cost for the creation of a pair but also for increasing the distance between particle and antiparticle.

Our study should be understood as a first step in the effort to simulate increasingly complex dynamics, including quantum simulations of lattice gauge theories<sup>5</sup>, that cannot be tackled by classical numerical methods. Building on these results, future challenges include the quantum simulation of non-Abelian lattice gauge theories and systems beyond one dimension.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 9 February; accepted 26 April 2016.**

1. Gattringer, C. & Lang, C. B. *Quantum Chromodynamics on the Lattice* (Springer, 2010).
2. Calzetta, E. A. & Hu, B. L. *Nonequilibrium Quantum Field Theory* (Cambridge Univ. Press, 2008).
3. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
4. Georgescu, I. M., Ashhab, S. & Nori, F. Quantum simulation. *Rev. Mod. Phys.* **86**, 153–185 (2014).
5. Jordan, S. P., Lee, K. S. M. & Preskill, J. Quantum algorithms for quantum field theories. *Science* **336**, 1130–1133 (2012).
6. Wiese, U. J. Ultracold quantum gases and lattice systems: quantum simulation of lattice gauge theories. *Ann. Phys.* **525**, 777–796 (2013).
7. Zohar, E., Cirac, I. & Reznik, B. Quantum simulations of lattice gauge theories using ultracold atoms in optical lattices. *Rep. Prog. Phys.* **79**, 014401 (2016).
8. Schwinger, J. Gauge invariance and mass. II. *Phys. Rev.* **128**, 2425–2429 (1962).

9. Kogut, J. & Susskind, L. Hamiltonian formulation of Wilson's lattice gauge theories. *Phys. Rev. D* **11**, 395–408 (1975).
10. Hebenstreit, F., Berges, J. & Gelfand, D. Real-time dynamics of string breaking. *Phys. Rev. Lett.* **111**, 201601 (2013).
11. Kasper, V., Hebenstreit, F., Oberthaler, M. & Berges, J. Schwinger pair production with ultracold atoms. Preprint at <http://www.arXiv.org/abs/1506.01238> (2015).
12. Hamer, C. J., Weihong, Z. & Oitmaa, J. Series expansions for the massive Schwinger model in Hamiltonian lattice theory. *Phys. Rev. D* **56**, 55–67 (1997).
13. Blatt, R. & Roos, C. F. Quantum simulations with trapped ions. *Nat. Phys.* **8**, 277–284 (2012).
14. Ladd, T. D. *et al.* Quantum computers. *Nature* **464**, 45–53 (2010).
15. Monroe, C. *et al.* Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects. *Phys. Rev. A* **89**, 022317 (2014).
16. Lanyon, B. P. *et al.* Universal digital quantum simulation with trapped ions. *Science* **334**, 57–61 (2011).
17. Lloyd, S. Universal quantum simulators. *Science* **273**, 1073–1078 (1996).
18. Narozhny, N. B. & Fedotov, A. M. Creation of electron-positron plasma with superstrong laser field. *Eur. Phys. J. Spec. Top.* **223**, 1083–1092 (2014).
19. Barends, R. *et al.* Digital quantum simulation of fermionic models with a superconducting circuit. *Nat. Commun.* **6**, 7654 (2015).
20. Salathé, Y. *et al.* Digital quantum simulation of spin models with circuit quantum electrodynamics. *Phys. Rev. X* **5**, 021027 (2015).
21. Zohar, E., Cirac, J. I. & Reznik, B. Simulating compact quantum electrodynamics with ultracold atoms: probing confinement and nonperturbative effects. *Phys. Rev. Lett.* **109**, 125302 (2012).
22. Tagliacozzo, L., Celi, A., Zamora, A. & Lewenstein, M. Optical Abelian lattice gauge theories. *Ann. Phys.* **330**, 160–191 (2013).
23. Banerjee, D. *et al.* Atomic quantum simulation of dynamical gauge fields coupled to fermionic matter: from string breaking to evolution after a quench. *Phys. Rev. Lett.* **109**, 175302 (2012).
24. Hauke, P., Marcos, D., Dalmonde, M. & Zoller, P. Quantum simulation of a lattice Schwinger model in a chain of trapped ions. *Phys. Rev. X* **3**, 041018 (2013).
25. Bañuls, M., Cichy, K., Cirac, I. & Jansen, K. The mass spectrum of the Schwinger model with matrix product states. *J. High Energy Phys.* **2013**, 158 (2013).
26. Sørensen, A. & Mølmer, K. Entanglement and quantum computation with ions in thermal motion. *Phys. Rev. A* **62**, 022311 (2000).
27. Schindler, P. *et al.* A quantum information processor with trapped ions. *New J. Phys.* **15**, 123012 (2013).
28. Gorin, T., Prosen, T., Seligman, T. H. & Znidaric, M. Dynamics of Loschmidt echoes and fidelity decay. *Phys. Rep.* **435**, 33–156 (2006).
29. Heyl, M., Polkovnikov, A. & Kehrein, S. Dynamical quantum phase transitions in the transverse field Ising model. *Phys. Rev. Lett.* **110**, 135704 (2013).
30. Plenio, M. B. & Virmani, S. An introduction to entanglement measures. *Quantum Inf. Comput.* **7**, 1–51 (2007).

**Acknowledgements** We acknowledge discussions with C. Hempel, E. R. Ortega and U. J. Wiese. Financial support was provided by the Austrian Science Fund (FWF), through the SFB FoQuS (FWF project nos F4002-N16 and F4016-N23), by the European Commission via the integrated project SIQS and the ERC synergy grant UQUAM, by the Deutsche Akademie der Naturforscher Leopoldina (grant nos LPDS 2013-07 and LPDR 2015-01), as well as the Institut für Quantenoptik und Quanteninformation GmbH. E.A.M. is a recipient of a DOC fellowship from the Austrian Academy of Sciences. P.S. was supported by the Austrian Science Foundation (FWF) Erwin Schrödinger Stipendium 3600-N27. This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Office grant W911NF-10-1-0284. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the US Government.

**Author Contributions** E.A.M., C.A.M., M.D. and T.M. developed the research based on discussions with P.Z. and R.B.; E.A.M. and P.S. performed the experiments. E.A.M., C.A.M., P.S. and M.H. analysed the data and carried out numerical simulations. E.A.M., P.S., D.N., A.E. and T.M. contributed to the experimental setup. C.A.M., M.H., M.D., P.H. and P.Z. developed the theory. E.A.M., C.A.M., P.S., M.H., P.H., M.D., P.Z. and R.B. wrote the manuscript and provided revisions. All authors contributed to discussions of the results and the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.A.M. ([esteban.martinez@uibk.ac.at](mailto:esteban.martinez@uibk.ac.at)) or C.A.M. ([christine.muschik@oeaw.ac.at](mailto:christine.muschik@oeaw.ac.at)).

**Reviewer Information** Nature thanks C. Wunderlich, E. Zohar and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Encoding of the lattice Schwinger model.** Our starting point is the Kogut–Susskind Hamiltonian formulation of the lattice Schwinger model<sup>8,9</sup>, see equation (1) in the main text. This model describes one-component fermion fields  $\hat{\psi}_n$  that are located at lattice sites  $n$  and interact with gauge fields that are represented by the canonically commuting operators  $[\hat{\theta}_n, \hat{L}_n] = i\delta_{n,m}$  as illustrated in Fig. 2.  $\hat{\theta}_n$  and  $\hat{L}_n$  represent the vector potential and electromagnetic field on the link connecting sites  $n$  and  $n+1$ . The dynamics is contained by local conservation laws. Formally, these are described in terms of local symmetry generators  $\hat{G}_i = \hat{L}_n - \hat{L}_{n-1} - \hat{\psi}_n^\dagger \hat{\psi}_n + \frac{1}{2}[1 - (-1)^n]$ . Physical states are eigenstates of these generators  $\hat{G}_i|\Psi_{\text{physical}}\rangle = q_i|\Psi_{\text{physical}}\rangle$ , where the  $q_i$  are background charges. In the continuum limit, we recover the familiar form of the Gauss law  $\nabla E = \rho$ , where  $\rho$  is the total charge density. The Schwinger Hamiltonian  $\hat{H}_{\text{lat}}$  commutes with the local symmetry generators  $[\hat{H}_{\text{lat}}, \hat{G}_i]$  and does not mix eigenstates of  $\hat{G}_i$  with different static charge configurations. We are interested in the case  $q_i = 0$ , that is, in the zero charge sector with an equal number of particles and antiparticles. Our dynamics is therefore constrained by the Gauss law:

$$\hat{L}_n - \hat{L}_{n-1} = \hat{\psi}_n^\dagger \hat{\psi}_n - \frac{1}{2}[1 - (-1)^n] \quad (2)$$

Equation (2) can be understood by considering a fixed field operator  $\hat{L}_n$  and an adjacent spin  $\hat{\psi}_n$  to its right. As shown in Fig. 2a, spins in state  $|\uparrow\rangle$  ( $|\downarrow\rangle$ ) on an odd (even) lattice site indicate that this lattice site is in the vacuum state, that is, not occupied by a particle or antiparticle. Accordingly,  $\hat{L}_n = \hat{L}_{n-1}$ . Spins in the state  $|\uparrow\rangle$  on even lattice sites (corresponding to positrons) generate  $(+1)$  unit of electric flux to the right  $\hat{L}_n = \hat{L}_{n-1} + 1$ . Similarly, spins in the state  $|\downarrow\rangle$  on odd lattice sites (corresponding to electrons) lead to a decrease of one unit,  $\hat{L}_n = \hat{L}_{n-1} - 1$ . In order to cast the lattice Schwinger Hamiltonian given in equation (1) in the main text in the form of a spin model, the one-component fermion operators  $\hat{\psi}_n$  are mapped to Pauli spin operators by means of a Jordan–Wigner transformation<sup>31</sup>:

$$\hat{\psi}_n = \prod_{l < n} [i\hat{\sigma}_l^z] \hat{\sigma}_n^-, \quad \hat{\psi}_n^\dagger = \prod_{l < n} [-i\hat{\sigma}_l^z] \hat{\sigma}_n^+$$

This leads to

$$\begin{aligned} \hat{H}_{\text{spin}} = & w \sum_{n=1}^{N-1} [\hat{\sigma}_n^+ e^{i\hat{\theta}_n} \hat{\sigma}_{n+1}^- + \text{h.c.}] \\ & + \frac{m}{2} \sum_{n=1}^N (-1)^n \hat{\sigma}_n^z + J \sum_{n=1}^{N-1} \hat{L}_n^2 \end{aligned}$$

where constant terms (energy offsets) have been omitted. Using this expression, the gauge degrees of freedom are eliminated in a two-step procedure<sup>12</sup>. First, the operators  $\hat{\theta}_n$  are eliminated by a gauge transformation:

$$\hat{\sigma}_n^- \rightarrow \prod_{l < n} [e^{-i\hat{\theta}_l}] \hat{\sigma}_n^-$$

In a second step, the electric field operators  $\hat{L}_n$  are eliminated iteratively using the spin version of the Gauss law given in equation (2):

$$\hat{L}_n - \hat{L}_{n-1} = \frac{1}{2}[\hat{\sigma}_n^z + (-1)^n]$$

This yields the pure spin Hamiltonian which is realized in our simulation scheme:

$$\begin{aligned} \hat{H}_S = & \frac{m}{2} \sum_{n=1}^N (-1)^n \hat{\sigma}_n^z + w \sum_{n=1}^{N-1} [\hat{\sigma}_n^+ \hat{\sigma}_{n+1}^- + \text{h.c.}] \\ & + J \sum_{n=1}^{N-1} \left[ \epsilon_0 + \frac{1}{2} \sum_{m=1}^n [\hat{\sigma}_m^z + (-1)^m] \right]^2 \end{aligned} \quad (3)$$

The free parameter  $\epsilon_0$  corresponds to the boundary electric field on the link to the left of the first lattice site (see Fig. 2b, c). Throughout this paper we consider the case of zero background field, where  $\epsilon_0 = 0$ .

The gauge fields do not appear explicitly in this description. Instead, they effectively generate a non-local long-range interaction that corresponds to the Coulomb interaction between the simulated charged particles. So far, this encoding approach has only been employed as a tool for analytical or numerical calculations<sup>12,32,33</sup>. In contrast, we investigate here the use of this idea for a quantum simulation scheme, that is, the realization

of the Schwinger model in its encoded form in an actual physical system. This approach has the advantage that, by construction, the dynamics takes place in the physically allowed subspace where the Gauss law is obeyed. In typical proposals for the quantum simulation of lattice gauge theories, this is fulfilled only up to some energy scale, as it is typically imposed energetically or by exploiting mechanisms where imperfections due to gauge-variant terms are strongly suppressed<sup>6</sup>.

**Digital quantum simulation of the encoded Schwinger model.** We realize  $\hat{H}_S$  given in equation (3) by means of a digital quantum simulation scheme<sup>17</sup>, which will be described in detail elsewhere (C.A.M. *et al.*, manuscript in preparation). For convenience, we express the simulated Hamiltonian in the form

$$\hat{H}_S = \hat{H}_{zz} + \hat{H}_{\pm} + \hat{H}_z \quad (4)$$

where the three parts of the Hamiltonian correspond to the two different types of two-body couplings  $\hat{H}_{zz}$  and  $\hat{H}_{\pm}$ , as well as local terms  $\hat{H}_z$ :

$$\hat{H}_{zz} = J \sum_{n < m} c_{nm} \hat{\sigma}_n^z \hat{\sigma}_m^z$$

$$\hat{H}_{\pm} = w \sum_n (\hat{\sigma}_n^+ \hat{\sigma}_{n+1}^- + \hat{\sigma}_{n+1}^+ \hat{\sigma}_n^-)$$

$$\hat{H}_z = m \sum_n c_n \hat{\sigma}_n^z + J \sum_n \tilde{c}_n \hat{\sigma}_n^z$$

The simulation protocol is based on time-coarse graining, where the desired dynamics of the Hamiltonian given by equation (3) is obtained within a time-averaged description. As illustrated in Fig. 2f, the total simulation time  $t_{\text{sim}}$  is divided into individual time windows of duration  $T$ . During each of these time windows, a full cycle of the protocol that is described below is performed. This cycle is repeated multiple times from  $t = 0$  to  $t = t_{\text{sim}}$  and consists of three sections, as shown in Fig. 2g. Each of these sections corresponds to one of the three parts of the desired Hamiltonian given by equation (4). In the first section,  $\hat{H}_{zz}$  is simulated, in the second, the nearest-neighbour terms  $\hat{H}_{\pm}$  are realized and in the third, the single particle rotations  $\hat{H}_z$  are performed. In this way, the simulation scheme uses only two types of interactions, local rotations and an infinite-range entangling operation

$$\hat{H}_{\text{MSx}} = J_0 \sum_{n,m} \hat{\sigma}_n^x \hat{\sigma}_m^x \quad (5)$$

which is routinely implemented in trapped ions by means of MS gates<sup>26</sup>. In the following, we explain how the individual parts of the Hamiltonian are realized. More detailed explanations can be found elsewhere (C.A.M. *et al.*, manuscript in preparation). The relative strengths of the individual parts of  $\hat{H}_S$ ,  $J$ ,  $w$  and  $m$ , can be tuned by adjusting the length of the elementary time windows or the strength of the underlying interaction  $J_0$  accordingly.

**Long-range interactions  $\hat{H}_{zz}$ .** The first part of equation (4) originates from the third term in equation (3) representing the electric-field energy. It takes the form

$$\hat{H}_{zz} = \frac{J}{2} \sum_{m=1}^{N-2} \sum_{n=m+1}^{N-1} (N-n) \hat{\sigma}_m^z \hat{\sigma}_n^z \quad (6)$$

and describes two-body interactions with an asymmetric distance dependence, where each spin interacts with constant strength with all spins to its left, while the coupling to the spins on its right decreases linearly with distance (see Fig. 2d, e). As the number of elements in the spin coupling matrix is proportional to  $N^2$ , a brute force digital simulation approach to this problem would require  $N^2$  time steps. Using our protocol, which is inspired by techniques put forward in ref. 34, the required resources scale only linearly in  $N$ . This is accomplished using the scheme illustrated in Fig. 2h. We introduce  $N-2$  time windows, which can be shown to be the minimal number of time steps required to simulate the Hamiltonian in equation (6). Each elementary time window has length  $\Delta t_i$ . In the  $n$ th time window, the Hamiltonian

$$\hat{H}_{\text{MSz}}^{(n)} = J_0 \sum_{i,j}^{n+1} \hat{\sigma}_i^z \hat{\sigma}_j^z$$

is applied.  $\hat{H}_{\text{MSz}}^{(n)}$  is realized by applying the Hamiltonian given in equation (5) in combination with local rotations,  $R(y) \hat{H}_{\text{MSz}} R^\dagger(y) = \hat{H}_{\text{MSz}}$ , where  $R(y) = e^{i\frac{\pi}{4} \sum_{i=1}^N \hat{\sigma}_i^y}$ . The resulting time-averaged Hamiltonian for the first section of the time interval  $T$ ,  $\bar{H}_I = \frac{1}{N-2} \sum_{n=1}^{N-2} \hat{H}_{\text{MSz}}^{(n)}$  is proportional to the desired Hamiltonian in equation (6),  $\bar{H}_I = \frac{J_0}{N-2} \hat{H}_{zz}$ .

As shown in Fig. 2h, only ions 1 to  $n+1$  participate in the entangling interaction in time step  $n$ . Since the interaction is implemented via a global beam that couples to the entire ion string (see Fig. 1b), ions  $n+2$  to  $N$  are decoupled by applying hiding pulses. The population in the qubit states of these ions is transferred to electronic levels that are not affected by the interaction using suitable laser pulses. The population in the state  $|\downarrow\rangle = 4S_{1/2}$  (magnetic number  $m = -1/2$ ) is transferred to the state  $3D_{5/2}$  ( $m = -5/2$ ), and the population in  $|\uparrow\rangle = 3D_{5/2}$  ( $m = -1/2$ ) is transferred to the state  $3D_{5/2}$  ( $m = -3/2$ ) via  $4S_{1/2}$  ( $m = +1/2$ ). *Nearest-neighbour terms*  $\hat{H}_{\pm}$ . The second part of equation (4),

$$\hat{H}_{\pm} = w \sum_{n=1}^{N-1} (\hat{\sigma}_n^+ \hat{\sigma}_{n+1}^- + \text{h.c.})$$

corresponds to the creation and annihilation of particle–antiparticle pairs (see Fig. 2a, c). For realizing this Hamiltonian, the interaction given in equation (5) needs to be modified not only in range, but also regarding the type of coupling. This is accomplished by dividing the time window dedicated to realizing  $\hat{H}_{\pm}$  (see Fig. 2g) into  $N-1$  elementary time slots of length  $\Delta t_{II}$ . Each of these is used for inducing the required type of interaction between a specific pair of neighbouring ions. For example, the first elementary time slot of length  $\Delta t_{II}$  is used to engineer an interaction of the type  $\hat{H}_{ij} \propto \hat{\sigma}_i^+ \hat{\sigma}_j^- + \text{h.c.}$  between the first and the second spin, the second time slot is used to do the same for the second and the third spin, and so on. This can be done by applying suitable hiding pulses to all spins except for a selected pair of ions  $i$  and  $j$ . The selected pair undergoes a sequence of gates, which transforms the  $\hat{\sigma}_i^x \hat{\sigma}_j^x$ -type coupling in equation (5) into an interaction of the required form and consists of four steps: (i) a single qubit operation on the two selected spins  $i$  and  $j$ ,  $U = e^{i\frac{\pi}{4}(\hat{\sigma}_i^z + \hat{\sigma}_j^z)}$ ; (ii) an evolution under the Hamiltonian given in equation (5) for the selected pair of spins,  $\hat{H}_{\text{MSX}}^{(ij)}$  during a time  $\Delta t_{II}/2$ ,  $e^{-i\hat{H}_{\text{MSX}}^{(ij)}\Delta t_{II}/2}$ ; (iii) another single qubit operation  $U^\dagger$ ; and finally (iv) another two-qubit gate  $e^{i\hat{H}_{\text{MSX}}^{(ij)}\Delta t_{II}/2}$ . The time evolution operator associated with the described sequence of gates is given by  $e^{i\hat{H}_{II}^{(ij)}\Delta t_{II}}$  with

$$\hat{H}_{II}^{(ij)} = \frac{1}{2} \left( \hat{H}_{\text{MSX}}^{(ij)} + U^\dagger \hat{H}_{\text{MSX}}^{(ij)} U \right) = J_0 (\hat{\sigma}_i^+ \hat{\sigma}_j^- + \text{h.c.})$$

as desired. The relative strength of the nearest-neighbour terms  $\hat{H}_{\pm}$  and the long-range couplings  $\hat{H}_{zz}$ ,  $w/J$  can be adjusted by tuning the ratio of the lengths of the elementary time windows  $\Delta_{II}/\Delta_I$ .

*Single-particle terms*  $\hat{H}_z$ . The last contribution to the Hamiltonian in equation (4) consists of two terms  $\hat{H}_z = m \sum_n c_n \hat{\sigma}_n^z + J \sum_n \tilde{c}_n \hat{\sigma}_n^z$ . The first term in this expression reflects the rest masses of the fermions. The second term is an effective single-particle contribution originating from the third part of equation (3) and corresponds to a change in the effective fermion masses due to the elimination of the electric fields. The local terms of the simulated Hamiltonian are given by:

$$\hat{H}_z = \frac{m}{2} \sum_{n=1}^N (-1)^n \hat{\sigma}_n^z - \frac{J}{2} \sum_{n=1}^{N-1} (n \bmod 2) \sum_{l=1}^n \hat{\sigma}_l^z$$

These are implemented by means of AC-Stark shifts, induced by laser pulses that are far red-detuned from the qubit transition<sup>13,27</sup>.

**Measurement and postselection.** For each set of system parameters and number of simulation time steps, we perform a full state tomography to determine the density matrix that corresponds to the quantum state of the system. The electronic state of the ions is detected via a fluorescence measurement using the electron shelving technique<sup>27</sup>. The entire string is imaged by a CCD camera, performing a full projective measurement in the  $z$  basis. This procedure is repeated 100 times to gather sufficient statistics.

As a consequence of charge conservation, an equal number of particles and antiparticles is created during the ideal dynamics of the system. Since our evolution starts with the vacuum state, the physical Hilbert space of the simulation is spanned by the six states  $\{|0000\rangle = |\uparrow\uparrow\downarrow\downarrow\rangle, |e^-e^+00\rangle = |\uparrow\uparrow\uparrow\downarrow\rangle, |0e^+e^-0\rangle = |\uparrow\uparrow\downarrow\downarrow\rangle, |00e^-e^+\rangle = |\uparrow\downarrow\downarrow\uparrow\rangle, |e^-00e^+\rangle = |\downarrow\downarrow\uparrow\uparrow\rangle, \text{ and } |e^-e^+e^-e^+\rangle = |\downarrow\downarrow\uparrow\uparrow\rangle\}$ , where  $|0\rangle$  denotes the vacuum,  $|e^- \rangle$  a particle and  $|e^+ \rangle$  an antiparticle. However, experimental errors during the simulation produce leakage from this subspace, such that non-physical states such as  $|e^-000\rangle = |\downarrow\downarrow\uparrow\uparrow\rangle$  get populated. Therefore, the raw measured density matrices  $\rho_{\text{raw}}$  are projected onto the Hilbert space spanned by the physical states and normalized,

$$\rho_{\text{phys}} = \frac{P \rho_{\text{raw}} P}{\text{tr}(P \rho_{\text{raw}} P)}$$

where  $P$  is the projector onto the physical subspace. All experimental data presented in this work correspond to physical density matrices  $\rho_{\text{phys}}$  postselected in

this way. The populations remaining in the physical subspace along the evolution are discussed in the following section.

**Experimental errors.** The bulk of the quantum gates in the simulation consists of hiding/unhiding pulses and MS gates. Each  $\pi$  pulse on a hiding transition has a fidelity of around 99.5%, and there are 30 such pulses per step, yielding a lower bound on the fidelity per step of  $(0.995)^{30} = 0.86$ . The fidelity of a fully-entangling ( $\pi/2$ ) MS gate on 4 ions is around 97.5%, and one simulation step has 8 quarter-entangling ( $\pi/8$ ) gates, yielding a lower bound of  $(0.975)^{8/4} = 0.95$ . The total lower bound for the fidelity per step is  $\mathcal{F} = (0.995)^{30} \cdot (0.975)^{8/4} = 76\%$ ; it is indeed lower than the average fidelity of the raw (not postselected) state after the first step, which is 89%. The sequence performs better than might be expected from the raw fidelities; we believe this is because the ideal evolution stays at all times in a decoherence-free subspace.

A useful measure of the performance of the evolution is the population leakage from the physical subspace. After  $\{1, 2, 3, 4\}$  evolution time steps, the measured populations remaining in the physical subspace were on average  $\{86 \pm 2, 79 \pm 1, 73 \pm 1, 69 \pm 1\}\%$  of the populations before postselection (the average is taken over the 7 simulation runs shown in the paper). The population loss per simulation step seems consistent with the errors induced by the hiding/unhiding operations.

The remaining errors can be quantified by the average fidelity of the postselected state with the ideal state. After the first evolution step this is 96%, which is consistent with the total fidelity of the MS gates. To quantify the performance of the simulation along the whole evolution, we compare the experimental data to a simple phenomenological error model. Since the postselection already partially corrects for population errors, we considered an error model that consists of uncorrelated dephasing, parameterized with an phase flip error probability  $p$  per qubit and per evolution time step. The density matrix  $\rho$  is then, at each evolution step, subject to the composition (denoted  $\circ$ ) of the error channels  $\mathcal{E}_i$  for each qubit

$$\rho \rightarrow \mathcal{E}_4 \circ \mathcal{E}_3 \circ \mathcal{E}_2 \circ \mathcal{E}_1(\rho)$$

where

$$\mathcal{E}_i(\rho) = (1-p)\rho + p\sigma_i^z \rho \sigma_i^z$$

The value for the error probability  $p$  was extracted from a fit to all of the experimental data collected. For all the data taken with non-zero  $J$  we found a value of  $p = 0.038$ . Whenever  $J = 0$ , the simulation does not require any  $zz$  interactions. Thus, several entangling gates are omitted from the sequence and consequently higher fidelities are expected. Indeed, for this case the error probability per time step was found to be  $p = 0.031$ .

**Quantum simulation of the Schwinger mechanism.** We simulate the coherent real-time dynamics in the Schwinger model focusing on the Schwinger mechanism, that is, spontaneous particle–antiparticle production out of the unstable vacuum. This effect is at the heart of quantum electrodynamics and its observation is currently pursued at high intensity laser facilities ELI and XCELS<sup>18</sup> (theoretical proposals for its quantum simulation can for example be found elsewhere<sup>6,7,35,36</sup>). To simulate the dynamics of pair creation, we consider as usual<sup>2,10</sup> the bare vacuum as initial state, where matter is completely absent,  $|\text{vacuum}\rangle = |0000\rangle$ . In the spin representation this state is accordingly given by  $|\uparrow\downarrow\uparrow\downarrow\rangle$ . Note that the bare vacuum is different from the so-called dressed vacuum state, which is the ground state of the full Hamiltonian.

*Decay of the vacuum.* The natural quantity characterizing the decay of the unstable vacuum is the vacuum persistence amplitude introduced by Schwinger<sup>37</sup>, which is defined as the overlap of the initial state  $|\Psi(0)\rangle = |\text{vacuum}\rangle$  with the time-evolved state

$$G(t) = \langle \text{vacuum} | e^{-i\hat{H}t} | \text{vacuum} \rangle$$

Within the original formulation, the Schwinger mechanism was considered for the continuum system and a classical electric field of strength  $E$  (ref. 37). There, it has been shown that the particle number density  $\nu(t)$  is directly related to the rate function  $\lambda(t)$  that characterizes the decay of the vacuum persistence probability  $|G(t)|^2$ :

$$\lambda(t) = - \lim_{N \rightarrow \infty} \frac{1}{N} \log[|G(t)|^2]$$

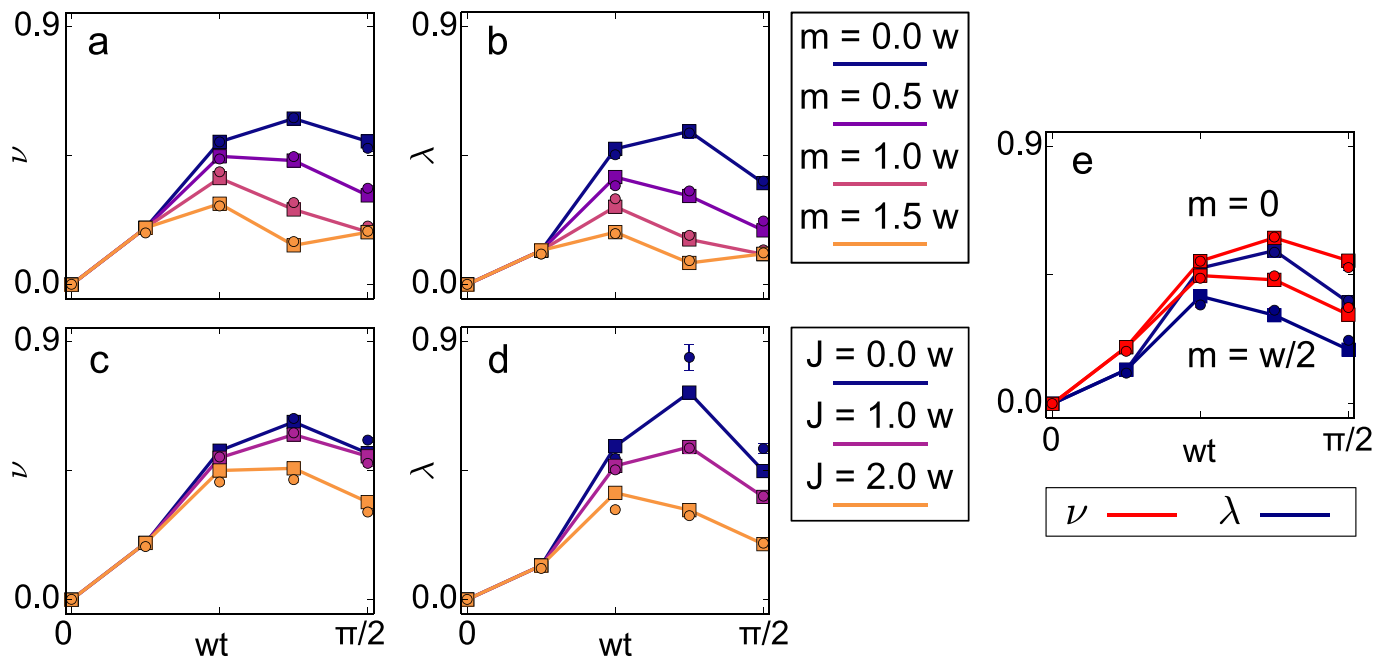
Specifically, in the limit of large fermion masses  $m \gg \sqrt{qE}$  with  $q$  the electric charge, as relevant in the high-energy context,  $\lambda(t) = \nu(t)$  for thermodynamically large systems in the continuum.

Since vacuum persistence amplitudes have so far not been measured, this connection between  $\lambda(t)$  and  $\nu(t)$  has not yet been tested experimentally. In Extended Data Fig. 1, we show the measured rate function  $\lambda(t)$  and find good qualitative agreement with  $\nu(t)$ , even for the few qubits in our digital quantum simulation.

*Finite size effects.* In the following, we discuss the dependence of the results on the number of lattice sites  $N$ . Extended Data Fig. 2 shows the time evolution of the

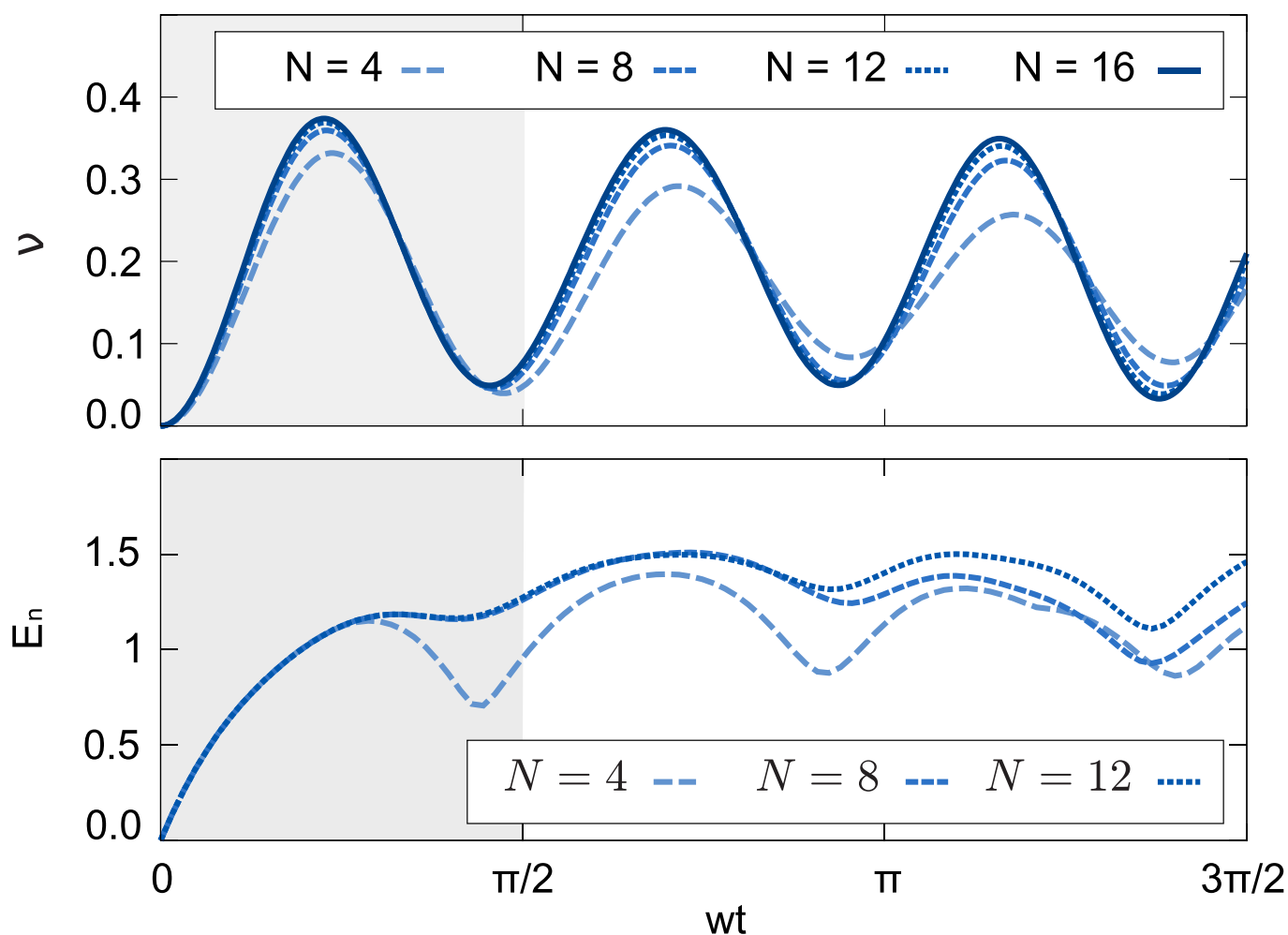
particle number density and the entanglement for different system sizes  $N$ . For our experimental system with  $N=4$ , we already find qualitative agreement with respect to the results expected for larger  $N$ . By scaling up the system, the dynamics quickly converges for the considered parameters. We address elsewhere the continuum limit  $a \rightarrow 0$ ,  $N \rightarrow \infty$  for fixed values of the coupling  $g$  and the mass  $m$  (C.A.M. *et al.*, manuscript in preparation).

31. Banks, T., Susskind, L. & Kogut, J. Strong-coupling calculations of lattice gauge theories:  $(1+1)$ -dimensional exercises. *Phys. Rev. D* **13**, 1043–1053 (1976).
32. Bañuls, M. C., Cichy, K., Cirac, J. I. & Jansen, K. The mass spectrum of the Schwinger model with matrix product states. *J. High Energy Phys.* **2013**, 158 [http://dx.doi.org/10.1007/JHEP11\(2013\)158](http://dx.doi.org/10.1007/JHEP11(2013)158) (2013).
33. Bañuls, M. C., Cichy, K., Cirac, J. I., Jansen, K. & Saito, H. Matrix product states for lattice field theories. *Proc. 31<sup>st</sup> Int. Symp. on Lattice Field Theory* 332 [http://pos.sissa.it/archive/conferences/187/332/LATTICE%202013\\_332.pdf](http://pos.sissa.it/archive/conferences/187/332/LATTICE%202013_332.pdf) (2013).
34. Graß, T., Muschik, C., Celi, A., Chhajlany, R. W. & Lewenstein, M. Synthetic magnetic fluxes and topological order in one-dimensional spin systems. *Phys. Rev. A* **91**, 063612 (2015).
35. Casanova, J. *et al.* Quantum simulation of quantum field theories in trapped ions. *Phys. Rev. Lett.* **107**, 260501 (2011).
36. García-Álvarez, L. *et al.* Fermion-fermion scattering in quantum field theory with superconducting circuits. *Phys. Rev. Lett.* **114**, 070502 (2015).
37. Schwinger, J. On gauge invariance and vacuum polarization. *Phys. Rev.* **82**, 664–679 (1951).



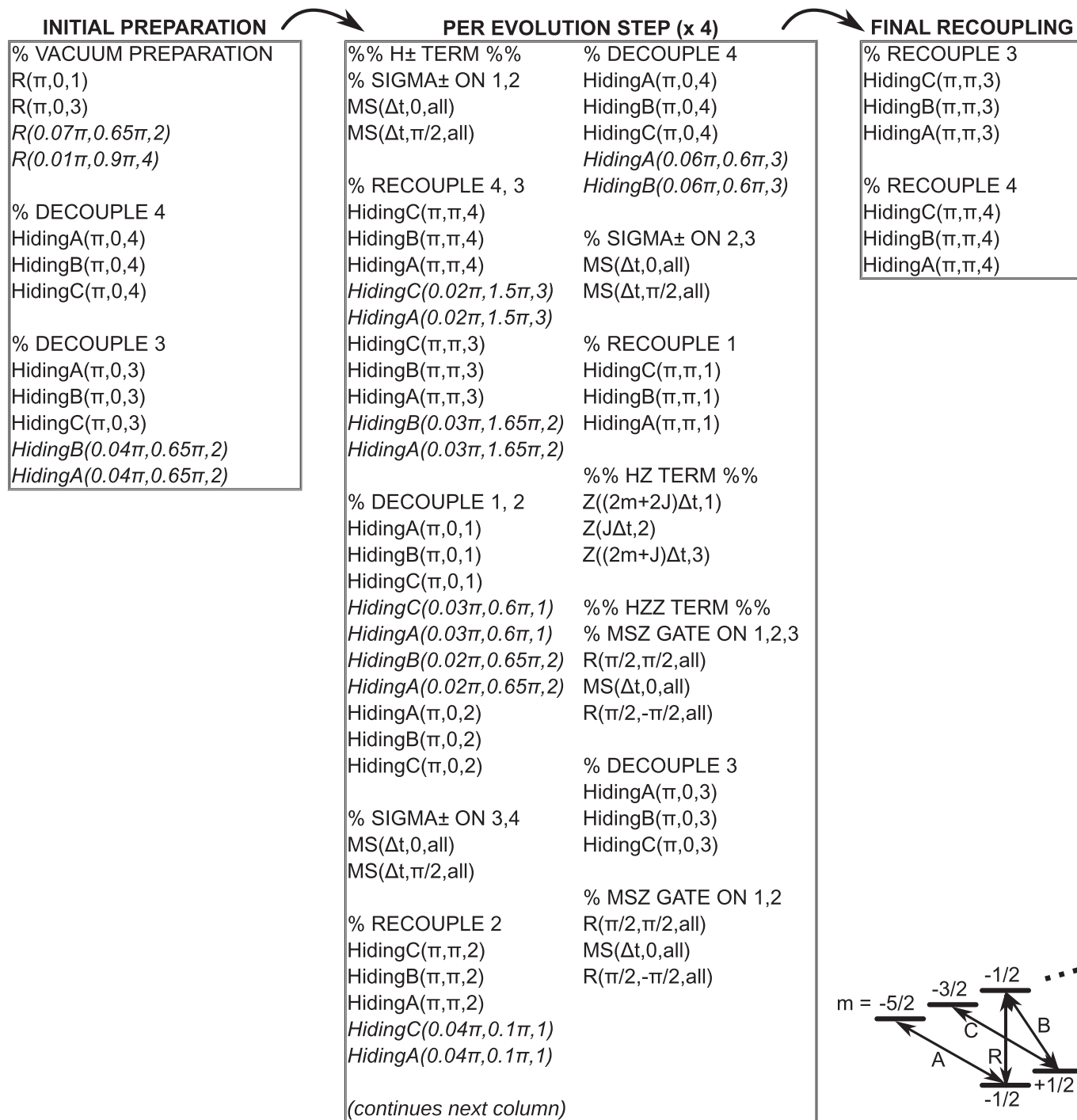
**Extended Data Figure 1 | Comparison of the evolutions of the particle number density  $\nu(t)$  and the rate function  $\lambda(t)$ .** The decay of the vacuum persistence probability is characterized by the rate function  $\lambda(t)$ , defined by  $|G(t)|^2 = e^{-N\lambda(t)}$ . **a, b**, Time evolution of  $\nu(t)$  (**a**) and  $\lambda(t)$  (**b**) for different values of the particle mass  $m$  and fixed electric field energy  $J=w$ , where  $w$  is the rate of particle–antiparticle creation and annihilation

(see equation (1) in the main text). **c, d**, Evolution of  $\nu(t)$  (**c**) and  $\lambda(t)$  (**d**) for different values of  $J$  and fixed particle mass  $m=0$  as a function of the dimensionless time  $wt$ . **e**, Comparison the evolutions of  $\nu(t)$  and  $\lambda(t)$  for  $J=w$  and masses  $m=0$  (upper two curves) and  $m=w/2$  (lower two curves). Error bars correspond to standard deviations estimated from a Monte Carlo bootstrapping procedure.



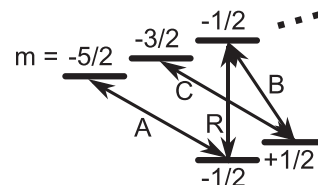
**Extended Data Figure 2 | Finite size effects.** Evolution of the particle number density  $\nu = \frac{1}{2N} \sum_{l=1}^N \langle (-1)^l \hat{\sigma}_l^z(t) + 1 \rangle$  (top) and the logarithmic negativity  $E_n$  (bottom) for different system sizes  $N$ . The logarithmic negativity is evaluated with respect to a cut in the middle of the considered

spin chain and quantifies the entanglement between the two halves of the system. Both quantities are shown as a function of the dimensionless time  $wt$  for  $J = m = w$ . The shaded area corresponds to the time interval explored in the experiment.



**Extended Data Figure 3 | Experimental pulse sequence.** This laser pulse sequence implements the evolution described in Fig. 2f, g. The pulses are listed in the order in which they are applied, as indicated by the arrows. The pulses in the first box prepare the initial state, those in the second box implement one step of the time evolution, and those in the third box recouple the ions to the computational subspace, that is, bring back their populations to the qubit transition  $4S_{1/2}(m = -1/2)$  to  $3D_{5/2}(m = -1/2)$ . The operations shown in the middle box are repeated once per evolution step, resulting in a total number of  $12 + 51 \times 4 + 6 = 222$  pulses for 4 evolution steps. The pulses are labelled in the form Pulse( $\theta, \phi$ , target

qubit), where  $\theta$  is the rotation angle (length) of the pulse,  $\phi$  its phase, and the target qubit is an integer from 1 to 4 for addressed operations or 'all' for global operations. 'R' denotes a pulse on the qubit transition  $4S_{1/2}(m = -1/2)$  to  $3D_{5/2}(m = -1/2)$ . 'MS' corresponds to an MS gate on the same transition. The hiding pulses 'HidingA,B,C' are applied on the transitions as follows: A,  $4S_{1/2}(m = -1/2)$  to  $3D_{5/2}(m = -5/2)$ ; B,  $4S_{1/2}(m = +1/2)$  to  $3D_{5/2}(m = -1/2)$ ; C,  $4S_{1/2}(m = +1/2)$  to  $3D_{5/2}(m = -3/2)$ . These transitions are shown in the level scheme at the bottom right. The pulses shown in italics serve the purpose of correcting addressing crosstalk.



# Solid-state harmonics beyond the atomic limit

Georges Ndashimiye<sup>1,2</sup>, Shambhu Ghimire<sup>2</sup>, Mengxi Wu<sup>3</sup>, Dana A. Browne<sup>3</sup>, Kenneth J. Schafer<sup>3</sup>, Mette B. Gaarde<sup>3</sup> & David A. Reis<sup>1,2</sup>

**Strong-field laser excitation of solids can produce extremely nonlinear electronic and optical behaviour. As recently demonstrated, this includes the generation of high harmonics extending into the vacuum-ultraviolet and extreme-ultraviolet regions of the electromagnetic spectrum<sup>1–8</sup>. High harmonic generation is shown to occur fundamentally differently in solids and in dilute atomic gases<sup>1–6,9–13</sup>. How the microscopic mechanisms in the solid and the gas differ remains a topic of intense debate<sup>1–11,14–18</sup>. Here we report a direct comparison of high harmonic generation in the solid and gas phases of argon and krypton. Owing to the weak van der Waals interaction, rare (noble)-gas solids are a near-ideal medium in which to study the role of high density and periodicity in the generation process. We find that the high harmonic generation spectra from the rare-gas solids exhibit multiple plateaus extending well beyond the atomic limit of the corresponding gas-phase harmonics measured under similar conditions. The appearance of multiple plateaus indicates strong interband couplings involving multiple single-particle bands. We also compare the dependence of the solid and gas harmonic yield on laser ellipticity and find that they are similar, suggesting the importance of electron–hole recollision in these solids. This implies that gas-phase methods such as polarization gating for attosecond pulse generation and orbital tomography could be realized in solids.**

Following the initial discovery of nonperturbative high-harmonic generation in solids<sup>1</sup>, several experimental<sup>2–8</sup> and theoretical<sup>10,11,14,15,17</sup> investigations have aimed to understand its detailed microscopic mechanism. In particular, the roles of the high density, periodicity and bonding and how they relate to atomic high harmonic generation (HHG) remains elusive. One striking difference between the harmonics from solids and from gases is in the scaling of the high-energy cutoff. In experiments on several materials (ZnO, SiO<sub>2</sub> and GaSe) with pump wavelengths spanning the terahertz to the near-infrared regions of the spectrum, the cutoff was found to scale linearly with the electric field for the solid<sup>1,4,5</sup>, whereas it scales linearly with the intensity for dilute gases<sup>12,13</sup>. In addition, in ZnO the ellipticity dependence was observed to be much weaker than in atomic gases<sup>1,2,19,20</sup>. The field-dependence of the cutoff and weak ellipticity dependence is consistent with a semi-classical Bloch oscillation model for the nonlinear intraband acceleration of electrons that have tunnelled across the direct bandgap<sup>1,4,5,18</sup>. The extent of the cutoff, to well beyond the maximum Bloch frequency, further suggests that the process is sensitive to details of the band structure through interactions beyond nearest neighbours<sup>1,4</sup>. However, interband contributions could also lead to a cutoff that is linear in the applied field, and the relative roles of inter- and intraband currents remains a topic of intense debate<sup>1–11,14–18</sup>.

In the generalized recollision picture proposed by Vampa *et al.*<sup>14</sup>, electrons in the conduction band recombine with their associated holes in the valence band in a manner that can be described by semi-classical trajectories. In this model, the energy of interband transitions is constrained to be less than the maximum band separation. Wu *et al.* have proposed that higher-lying conduction bands

can give rise to multiple plateaus in the HHG spectrum, each with a cutoff that is limited only by the field-dressed energy spacing between bands<sup>10</sup>. Using a semiconductor Bloch equations treatment, Schubert *et al.* found that at far-infrared wavelengths, the HHG spectrum is dominated by intraband dynamics<sup>5</sup>.

Until now, experiments have concentrated on covalently<sup>1,4–8</sup> bonded crystals. In such crystals the overlap of the atomic and molecular wavefunctions leads to a strong modification of the electronic states, making it difficult to extract material-independent aspects of the strong-field process. Rare-gas solids (RGS) are the nearest thing to a three-dimensional array of isolated atoms at high density<sup>21</sup>, owing to the closed shell structure and high ionization potential of rare gases, and their weak bonding due to van der Waals interactions. Here we study the HHG from Ar and Kr in both gas and solid phases. We find that the HHG spectra from the solids exhibit multiple plateaus extending beyond both single-atom predictions and our measured gas harmonics for the same laser parameters. In addition, the photon energy of the solid-state harmonics greatly exceeds the maximum band separation between the highest-valence and lowest-conduction bands, suggesting the importance of solid-state effects and electronic band structure even in weakly bonded RGS.

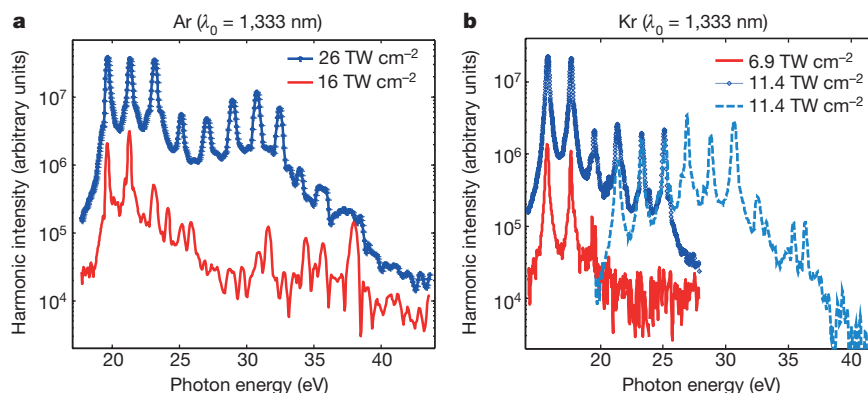
The experiments were performed on 5- $\mu\text{m}$ -thick polycrystalline Ar and Kr RGS and 3 Torr of Ar and Kr in a 1-mm-long cell for the gas (see Methods). The targets were irradiated by beams of 50 fs, 1,333 nm (0.93 eV) and 50 fs, 1,500 nm (0.82 eV) generated by a 40-fs, 1-kHz amplified Ti:sapphire laser-pumped optical parametric amplifier. The intensity of the infrared beam was calibrated using the measured spectral cutoff in the Ar and Kr gas, assuming that the cutoff energy is given by the microscopic value<sup>12,13</sup>,  $E_{\text{cut}} = I_p + 3.17U_p$ , where  $I_p$  is the ionization energy and  $U_p$  is the ponderomotive energy of a free electron in the laser field.

Figure 1 shows representative high harmonic spectra from solid Ar and Kr at two intensities each, using the 1,333-nm pump (16 TW cm<sup>−2</sup> and 26 TW cm<sup>−2</sup> for Ar in Fig. 1a and 6.9 TW cm<sup>−2</sup> and 11.4 TW cm<sup>−2</sup> for Kr in Fig. 1b). The spectra from Ar and Kr are qualitatively similar, although the laser intensities for Kr are about a factor of two lower than Ar. A single plateau is evident in each at the lower intensity. For the higher intensity a second plateau is evident at ~25–33 eV for Ar and at ~19–31 eV for Kr. Figure 2a shows the harmonic spectra as a function of intensity for Ar at the 1,333-nm pump. As can be seen here, the second plateau appears suddenly over a very narrow intensity range and comprises several harmonics. The presence of the second plateau is in contrast to the experiment on HHG in dilute atomic gases, including those measured here, where only a single plateau is observed. Also shown in Fig. 2b are theoretical calculations, described below.

The high-energy extent of the second plateau notably exceeds the cutoff for the gas for the same drive wavelength and intensity. At moderate peak intensities (10 TW cm<sup>−2</sup> for Kr and 20 TW cm<sup>−2</sup> for Ar), the secondary plateau is almost entirely beyond the projected gas-phase cutoff. Figure 3 shows the cutoff as a function of intensity for Ar and Kr RGS and rare gases for the two different pump wavelengths.

<sup>1</sup>Department of Applied Physics, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA.

<sup>3</sup>Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA.



**Figure 1 | Representative spectra of HHG from solid Ar and Kr on a logarithmic scale for the driving wavelength of  $\lambda_0 = 1,333$  nm.**

**a**, HHG spectra from solid Ar. The spectrum taken at low intensity has only harmonics of the first plateau. At higher intensity, there are harmonics in a second plateau that start from the 27th harmonic (25 eV) and end at the 35th harmonic (32.5 eV). A third plateau is present, but its harmonics are dimmer and only its first two harmonics can be

distinguished from the background. **b**, HHG spectrum from solid Kr.

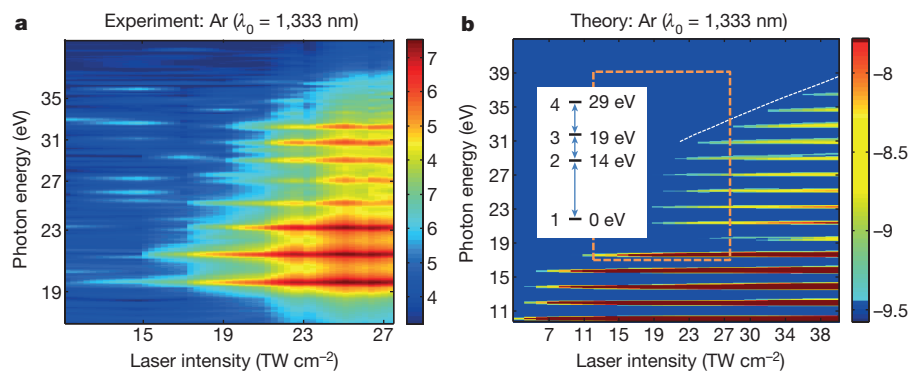
The spectra are taken at different spectrometer configurations and two spectra from different spectrometer configurations have been concatenated for the higher intensity. The spectra of solid Kr behave in a way similar to that of the spectrum of solid Ar except that for solid Kr the harmonics in the second plateau start at the 21st harmonic (19.5 eV) and end at the 37th harmonic (34.4 eV) and are much dimmer than in solid Ar.

The cutoff of the solid harmonics increases with increasing intensity in a nontrivial way, neither following the linear cutoff in intensity of the gas nor the square-root of intensity seen for other solids. The RGS cutoff is below the rare-gas cutoff at low intensity; however, with the sudden onset of a second plateau (as seen in Fig. 1) the RGS cutoff notably exceeds the rare-gas cutoff for the same drive wavelength and intensity. Even at moderate peak intensities ( $10 \text{ TW cm}^{-2}$  for Kr and  $20 \text{ TW cm}^{-2}$  for Ar) the second plateau is almost entirely beyond the projected gas phase cutoff. The maximum photon energy detected using both wavelengths exceeds the maximum separation between the highest-valence and the lowest-conduction bands. According to Bacalis *et al.*<sup>22</sup>, this separation is around 19 eV in solid Ar and 16 eV in solid Kr. In fact, the maximum band separation is below the higher end of the first plateau by about 3 eV in both solids.

The nontrivial scaling of the high-energy cutoff with intensity and the sudden appearance of multiple plateaus are indicative of complex solid-state behaviour. The appearance of multiple plateaus can be understood in a model in which solid-state HHG results from strong-field-driven transitions of Bloch electrons<sup>10,11</sup>. This behaviour is linked to the coupling of pairs of higher-lying conduction bands that are reached in a step-like process<sup>10</sup>. The cutoff energy and the strength of each plateau depend on the energy separation and the coupling

strength between pairs of bands<sup>23</sup>, and different plateaus can therefore exhibit different nonlinear scaling with laser intensity. Two general features of such multi-band couplings are the sudden appearance of the second plateau and the different slopes of cutoff energy versus intensity that are observed for the two plateaus.

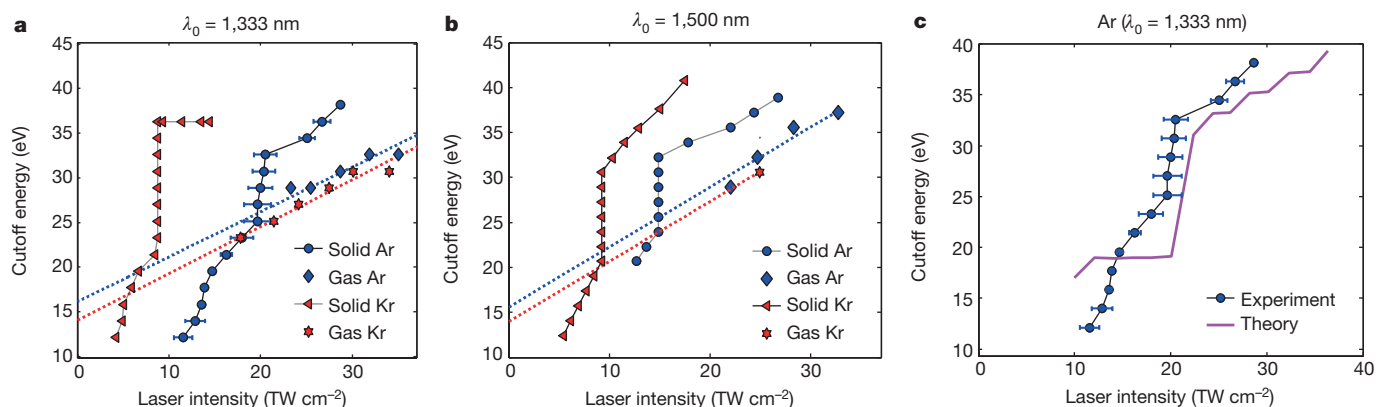
We apply this model for HHG in solid argon by solving the time-dependent Schrödinger equation for a four-band system in which the energy levels and dipole transition elements originate from a density functional theory (DFT)-based band structure calculation<sup>24</sup>. In ref. 10, we show the formal equivalence between taking a Houston-state basis, where the electron and hole wavevectors  $k$  are a function of the instantaneous vector potential of the laser, and a Bloch-state basis, where the electron transitions are between different bands at fixed  $k$ . Therefore, in this single-particle picture both the interband and the intraband contribution to the HHG yield will be included when considering discrete states at  $k = 0$ , assuming that the tunnelling is concentrated near the direct gap at the zone centre ( $\Gamma$  point). Figure 2b shows the calculated harmonic spectrum as a function of laser intensity. The energies for the relevant electronic states at  $\Gamma$  are shown in the inset. The first plateau originates in the coupling between levels 1 and 2, and the second plateau in the coupling between 2 and 3. The dashed white curve indicates the prediction



**Figure 2 | Evolution of the HHG spectrum as a function of the laser intensity.** The colour scale shows the logarithm of the intensity.

**a**, Experimental data. HHG of solid Ar using 1,333-nm drive laser. At moderate peak intensities, the high-energy cutoff increases smoothly (up to the 27th harmonic). At around  $20 \text{ TW cm}^{-2}$  the spectral cutoff increases suddenly to the 35th harmonic. The first plateau is brighter than the second at all intensities. **b**, Theoretical results. HHG spectrum obtained by solving the time-dependent Schrödinger equation for a four-level system in which the energy separation and couplings between

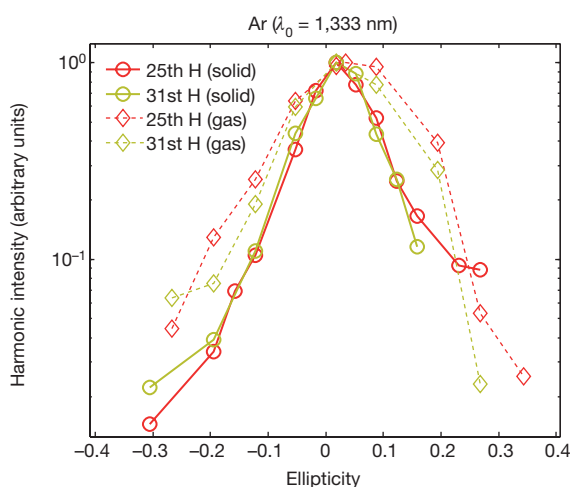
levels correspond to those of the Ar band-structure at the zone centre ( $\Gamma$  point). The calculation qualitatively and semiquantitatively reproduces the experimental data. The dashed rectangle shows the corresponding range of the experimental data. The inset shows the energy levels used in the calculations, with arrows representing the coupling between levels. The dashed white curve indicates the prediction for the cutoff energy of the second plateau based on the energy difference between field-dressed levels 1 and 3.



**Figure 3 | Comparison of the spectral cutoff of HHG in solid and gas Ar and Kr as a function of the laser intensity.** The dotted straight lines are fits of the linear cutoff of gas harmonic (blue for Ar and red for Kr). At zero laser intensity the fit was constrained to be at the ionization energies of the Ar and Kr gases ( $I_p = 15.7$  eV and 14 eV), respectively. **a**, Spectral cutoff of the harmonics of 1,333 nm. For solid Ar and Kr the cutoff is not linear in the electrical field or intensity below the harmonics of photon energy 25 eV or 20 eV, respectively. In these laser intensity regions the cutoff is below that of a corresponding gas HHG (Ar and Kr). Above these photon energies, the cutoff curves abruptly turn to an almost vertical slope which indicates a switching on of new HHG processes for both solid Ar and Kr. Above 33 eV for Ar the cutoff slope decreases to a value almost equal to the slope before the sudden rise (below 25 eV). However, for solid Kr after the sudden increase in the cutoff slope, which

started around 20 eV to around 35 eV, no higher harmonic was observed above 35 eV even with increasing laser intensity. The error bars are the standard deviation of repeated cutoff measurements of solid Ar HHG. **b**, Spectral cutoff of the harmonics of 1,500 nm. For both solid Ar and Kr, the HHG cutoff respectively below 25 eV and 20 eV is almost linear in intensity and this cutoff is below of that of the corresponding gases. Above these photon energies the slope of the cutoff curves becomes vertical, as for the 1,333-nm case. However, for solid Kr the sudden increase in the cutoff energy ends at 31 eV, and a less steep cutoff develops between 31 eV and 41 eV. **c**, Comparison of measured cutoff and calculated cutoff. There is qualitative agreement between the measurement and the calculation. The error bars are the standard deviation of repeated cutoff measurements of solid Ar HHG.

for the cutoff energy of the second plateau based on the energy difference between the field-dressed levels 1 and 3 (see Methods). The calculated values of the first and second cutoff energies compare reasonably well to those of the experiment, as shown in Fig. 2b and Fig. 3c. We note that to get the best agreement on the location and scaling of the plateaus to the experimental results, we have adjusted the dipole coupling strengths obtained from DFT (see Methods). In this case, the coupling strengths between levels 1 and 2, between levels 2 and 3, and between levels 3 and 4 are comparable. This is indicative of a periodic system in which electrons are strongly localized around the individual atomic sites rather than the more delocalized electron behaviour one would find in covalently bonded semiconductors such as ZnO.



**Figure 4 | Comparison of ellipticity dependence of the 25th and 31st harmonics from solid and gas-phase Ar at similar peak fields.** The 25th and 31st harmonics lay in the first and second plateaus, respectively, for solid Ar. The harmonic intensity shows a similar ellipticity dependence for the two orders from both the solid and gas. The sensitivity to ellipticity is at least as strong in the solid as in the gas.

This conclusion is supported by the measured strong ellipticity dependence in RGS compared to what was previously measured in ZnO<sup>1</sup>. In Fig. 4, we show our measurements of the ellipticity dependence for two representative harmonics of gas and solid Ar. These harmonics were chosen such that they fall within the ranges of the first and second plateaus for the RGS (25th and 31st harmonics, respectively). These harmonics show similar ellipticity dependence for the different orders. Moreover, the solid-state harmonics are at least as sensitive to ellipticity as is the gas. We note that the strong ellipticity dependence in atomic and molecular HHG has been attributed to the transverse momentum of the tunnel-ionized electron causing the returning electron wavepacket to miss the parent ion<sup>19,20</sup>. That characteristic has been used in the generation of isolated attosecond pulses<sup>25</sup> and in imaging of molecular orbitals<sup>26</sup>. Our observation of strong ellipticity dependence of harmonics in RGS opens up similar opportunities for attosecond pulse generation and imaging of electronic wavefunctions in the solid state.

We have measured high harmonics from solid and gas-phase Ar and Kr. The solid-state harmonics include multiple plateaus whose high-energy cutoff extends beyond the atomic limit. This demonstrates the importance of solid-state effects even in the weakly bound van der Waals solids. We show that the multiple plateaus can be accommodated in a single-particle picture, similar to the single-active electron models used to calculate gas-phase harmonics in both atomic and molecular systems, in which only one electron is assumed to interact with the laser field<sup>12,13</sup>. We further note that the appearance of the second plateau occurs at very nearly twice the exciton energy ( $\sim 24$  eV in Ar and  $\sim 20$  eV in Kr). Although this could be indicative of many body-effects beyond the single-particle picture, such a model is not required to capture the general features seen in the experiment. In either case, the important difference between the solid and the gas is that the solid involves transitions between bound states created by band folding due to the periodic potential. In this sense there is no free-particle continuum for electrons in the solid, or equivalently, even in the case of RGS, the effect of the periodic potential cannot be neglected.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 December 2015; accepted 4 March 2016.**

**Published online 6 June 2016.**

- Ghimire, S. *et al.* Observation of high-order harmonic generation in a bulk crystal. *Nature Phys.* **7**, 138–141 (2011).
- Ghimire, S. *et al.* Strong-field and attosecond physics in solids. *J. Phys. At. Mol. Opt. Phys.* **47**, 204030 (2014).
- Ghimire, S. *et al.* Redshift in the optical absorption of ZnO single crystals in the presence of an intense midinfrared laser field. *Phys. Rev. Lett.* **107**, 167407 (2011).
- Luu, T. T. *et al.* Extreme ultraviolet high-harmonic spectroscopy of solids. *Nature* **521**, 498–502 (2015).
- Schubert, O. *et al.* Sub-cycle control of terahertz high-harmonic generation by dynamical Bloch oscillations. *Nature Photon.* **8**, 119–123 (2014).
- Hohenleutner, M. *et al.* Real-time observation of interfering crystal electrons in high-harmonic generation. *Nature* **523**, 572–575 (2015).
- Vampa, G. *et al.* Linking high harmonics from gases and solids. *Nature* **522**, 462–464 (2015).
- Vampa, G. *et al.* All-optical reconstruction of crystal band structure. *Phys. Rev. Lett.* **115**, 193603 (2015).
- Ghimire, S. *et al.* Generation and propagation of high-order harmonics in crystals. *Phys. Rev. A* **85**, 043836 (2012).
- Wu, M., Ghimire, S., Reis, D. A., Schafer, K. J. & Gaarde, M. B. High-harmonic generation from Bloch electrons in solids. *Phys. Rev. A* **91**, 043839 (2015).
- Korbman, M., Yu Kruchinin, S. & Yakovlev, V. S. Quantum beats in the polarization response of a dielectric to intense few-cycle laser pulses. *New J. Phys.* **15**, 013006 (2013).
- Corkum, P. B. Plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Krause, J. L., Schafer, K. J. & Kulander, K. C. High-order harmonic generation from atoms and ions in the high intensity regime. *Phys. Rev. Lett.* **68**, 3535–3538 (1992).
- Vampa, G. *et al.* Theoretical analysis of high-harmonic generation in solids. *Phys. Rev. Lett.* **113**, 073901 (2014).
- McDonald, C. R., Vampa, G., Corkum, P. B. & Brabec, T. Interband Bloch oscillation mechanism for high-harmonic generation in semiconductor crystals. *Phys. Rev. A* **92**, 033845 (2015).
- Vampa, G., McDonald, C. R., Orlando, G., Corkum, P. B. & Brabec, T. Semiclassical analysis of high harmonic generation in bulk crystals. *Phys. Rev. B* **91**, 064302 (2015).
- Higuchi, T., Stockman, M. I. & Hommelhoff, P. Strong-Field Perspective on High-Harmonic Radiation from Bulk Solids. *Phys. Rev. Lett.* **113**, 213901 (2014).
- Golde, D., Meier, T. & Koch, S. W. Microscopic analysis of high-harmonic generation in semiconductor nanostructures. *Phys. Status Solidi C* **6**, 420–423 (2009).
- Budil, K. S., Salières, P., L'Huillier, A., Ditmire, T. & Perry, M. D. Influence of ellipticity on harmonic generation. *Phys. Rev. A* **48**, R3437–R3440 (1993).
- Burnett, N. H., Kan, C. & Corkum, P. B. Ellipticity and polarization effects in harmonic generation in ionizing neon. *Phys. Rev. A* **51**, R3418–R3421 (1995).
- Klein, M. L. & Venables, J. A. *Rare Gas Solids* 123 (Academic Press, 1976).
- Bacalis, N. C., Papaconstantopoulos, D. A. & Pickett, W. E. Systematic calculations of the band structures of the rare-gas crystals neon, argon, krypton, and xenon. *Phys. Rev. B* **38**, 6218–6226 (1988).
- Gauthey, F. I., Garraway, B. M. & Knight, P. L. High harmonic generation and periodic level crossings. *Phys. Rev. A* **56**, 3093–3096 (1997).
- Blaha, P., Schwarz, K., Madsen, G., Kvasnicka, D. & Luitz, J. *WIEN2k, An Augmented Plane Wave Plus + Local Orbitals Program for Calculating Crystal Properties* (TU Wien, 2001).
- Shan, B., Ghimire, S. & Chang, Z. Generation of the attosecond extreme ultraviolet supercontinuum by a polarization gating. *J. Mod. Opt.* **52**, 277–283 (2005).
- Itatani, J. *et al.* Tomographic imaging of molecular orbitals. *Nature* **432**, 867–871 (2004).

**Acknowledgements** At Stanford/SLAC, this work was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, through the AMOS programme within the Chemical Sciences Division (G.N., D.A.R.) and the Office of Science Early Career Research Program (S.G.). At Louisiana State University this work is supported by the National Science Foundation under grant number PHY-1403236. Solid Ar samples were characterized at Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, the use of which is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract number DE AC02-76SF00515.

**Author Contributions** G.N., S.G. and D.A.R. conceived the experiments. G.N. performed the experiment and analysed the data. M.W., K.J.S., and M.B.G. developed the single electron laser excitation theory and performed the calculation. D.A.B. performed the DFT calculations. All authors contributed to the interpretation of the results and writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A.R. ([dreis@stanford.edu](mailto:dreis@stanford.edu)) or G.N. ([ndabashi@stanford.edu](mailto:ndabashi@stanford.edu)).

## METHODS

**Experiments.** The RGS samples were grown inside an ultrahigh vacuum ( $<10^{-7}$  mbar) chamber at a temperature of 20 K for Ar and 27 K for Kr. A closed-circuit cryostat was used to cool down a silicon wafer with 30-nm-thick silicon nitride windows ( $0.5\text{ mm} \times 0.5\text{ mm}$ ). The silicon nitride was used as the substrate for growing the crystals. The Ar and Kr crystals were grown at a rate of about  $1\text{ }\mu\text{m min}^{-1}$  and the thickness was measured using thin-film interference of a HeNe laser beam. Growth occurred in the absence of the strong-field infrared pump beam. The silicon nitride windows were ablated at the focal spot by the strong-field pump subsequent to growth, leaving a free-standing RGS film. Sublimation of the samples was found to be negligible over the course of the exposure by the laser when excited below the damage threshold. The damage threshold was determined to be  $\sim 30\text{ TW cm}^{-2}$  for solid Ar and  $\sim 15\text{ TW cm}^{-2}$  for solid Kr by both visual inspection and loss of the harmonic emission and for the conditions reported here. The harmonic spectra were measured using a home-built spectrometer consisting of a flat-field imaging grating (Hitachi 001-0639) and micro channel plate (MCP) detection system. The MCP was mounted on movable bellows which permitted us to observe different spectral regions between 10 eV and 45 eV. The high energy cutoff was determined by the intensity at which the highest harmonic is approximately three times the noise on the MCP, ignoring peaks that are inconsistent across multiple measurements. We note that the intensities used here are about an order of magnitude lower than in the typical gas-phase HHG experiments, and therefore reaching the fully phase-matched conditions in the gas would require a much higher pressure than we can achieve with our experimental setup. Because of that technical limitation, we performed measurements at relatively high peak intensity ( $>20\text{ TW cm}^{-2}$  in Ar gas), and extrapolated the high-energy cutoff scaling results to the moderate peak intensity scales. The intercept on the cutoff energy axis corresponds to the ionization energy threshold, as expected. For the ellipticity dependence measurements, the polarization was varied using a quarter-wave plate. The peak field along the major axis was kept constant by adjusting the pulse energy.

The Ar films were characterized under the same growth conditions by X-ray diffraction at the Stanford Synchrotron Radiation Lightsource and the size of the crystal grains was found to be at least 100 nm. The sample thickness was chosen empirically to be thick enough to provide mechanical stability against the strong-field excitation and substrate ablation, but thin enough to mitigate propagation effects, including cascaded nonlinear wavemixing, which we observed to depend on film thickness<sup>27</sup>. The focal spot upon propagation through the solid film and divergence of the harmonics were measured to be independent of incident intensity, falling within 10% of their nominal values.

**Calculations.** The argon band structure and dipole transition elements are calculated using DFT. The DFT calculations employ the linear augmented plane-wave method implemented in Wien2k<sup>24</sup>. The calculations use a muffin-tin radius of 3.0 Bohr, a  $k$ -point grid of  $33 \times 33 \times 33$  in the Brillouin zone, and a plane-wave energy cutoff of 50 atomic units. The DFT code uses the Perdew–Burke–Ernzeroff generalized gradient approximation (GGA) functional<sup>28</sup>. Since DFT produces an energy gap that is too small, a modified Becke–Johnson correction<sup>29</sup> is applied to the

conduction-bands energies to obtain better agreement with the experimental band structure.

We next use the band structure and the transition matrix elements to solve the time-dependent Schrödinger equation (TDSE) in  $k$ -space, in which different  $k$ -values are decoupled. The harmonic spectrum is obtained as the Fourier transform of the time-dependent current calculated from the TDSE solution<sup>10</sup>. Our initial condition is a delocalized Bloch wavefunction (only  $k=0$ ) located at the highest symmetric point,  $\Gamma$ , on the valence band. For simplicity we include only the four lowest strongly coupled bands, meaning that we are solving the TDSE for a four-level system, in the Bloch basis<sup>10</sup>. We have established that including additional higher bands makes only a negligible change to the harmonic spectrum, whereas excluding any of the four lowest bands makes a large difference. We adjust the dipole transition matrix elements so that the couplings between levels 1 and 2, between levels 2 and 3, and between levels 3 and 4 are roughly equal (in atomic units the couplings are  $\mu_{1,2}=0.62$ ,  $\mu_{2,3}=0.41$ , and  $\mu_{3,4}=-0.61$ ) and ignore all other couplings. We justify this adjustment on the grounds that the experiment averages over all orientations, and precise dipole-matrix elements for higher bands are difficult to extract from DFT. The four strongly coupled levels in our model represent the simplest possible description of the coupling between clusters of bands, which would be expected to occur in steps, with approximately equal coupling strengths.

In the four-level picture, the multiple plateaus simply come from transitions between the instantaneous, field-dressed, eigenstates of the system. For a four-level system driven by a laser field  $F(t)$ , the Hamiltonian can be written as

$$H = \begin{pmatrix} \omega_1 & \mu_{12}F(t) & 0 & 0 \\ \mu_{12}F(t) & \omega_2 & \mu_{23}F(t) & 0 \\ 0 & \mu_{23}F(t) & \omega_3 & \mu_{34}F(t) \\ 0 & 0 & \mu_{34}F(t) & \omega_4 \end{pmatrix}$$

where  $\omega_i$  is the energy of the band  $i$  and  $\mu_{ij}$  is the dipole transition between levels  $i$  and  $j$ . The energies of the instantaneous eigenstates  $E_i$  can be calculated by diagonalizing the Hamiltonian. Then the cutoff energy for the first plateau will be given by the maximum energy difference between the field-dressed levels,  $(E_2 - E_1)_{\text{max}}$ , which scales approximately linearly with the laser field strength<sup>10</sup>. Likewise the cutoff energy for the second plateau will be  $(E_3 - E_1)_{\text{max}}$ . This prediction agrees very well with the two harmonic cutoffs visible in Fig. 2b.

27. Ndashimiye, G., Ghimire, S., Reis, D. & Nicholson, D. Measurement of coherence lengths of below threshold harmonics in solid argon. *Opt. Soc. Am. Tech. Dig.* QW1A.7 [http://dx.doi.org/10.1364/CLEO\\_QELS.2013.QW1A.7](http://dx.doi.org/10.1364/CLEO_QELS.2013.QW1A.7) (2013).
28. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
29. Tran, F. & Blaha, P. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.* **102**, 226401 (2009).

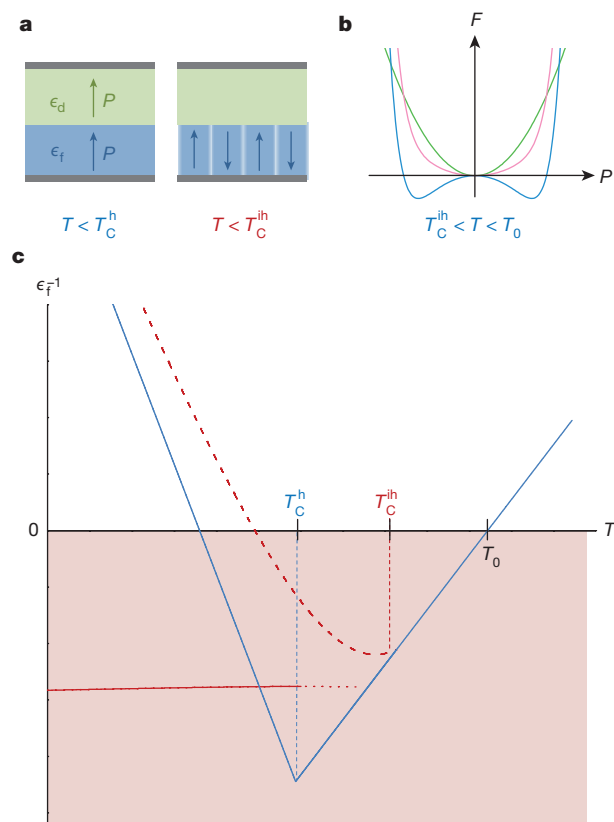
# Negative capacitance in multidomain ferroelectric superlattices

Pavlo Zubko<sup>1\*</sup>, Jacek C. Wojdel<sup>2\*</sup>, Marios Hadjimichael<sup>1</sup>, Stéphanie Fernandez-Pena<sup>3</sup>, Anaïs Sené<sup>4</sup>, Igor Luk'yanchuk<sup>4,5</sup>, Jean-Marc Triscone<sup>3</sup> & Jorge Íñiguez<sup>2,6</sup>

The stability of spontaneous electrical polarization in ferroelectrics is fundamental to many of their current applications, which range from the simple electric cigarette lighter to non-volatile random access memories<sup>1</sup>. Research on nanoscale ferroelectrics reveals that their behaviour is profoundly different from that in bulk ferroelectrics, which could lead to new phenomena with potential for future devices<sup>2–4</sup>. As ferroelectrics become thinner, maintaining a stable polarization becomes increasingly challenging. On the other hand, intentionally destabilizing this polarization can cause the effective electric permittivity of a ferroelectric to become negative<sup>5</sup>, enabling it to behave as a negative capacitance when integrated in a heterostructure. Negative capacitance has been proposed as a way of overcoming fundamental limitations on the power consumption of field-effect transistors<sup>6</sup>. However, experimental demonstrations of this phenomenon remain contentious<sup>7</sup>. The prevalent interpretations based on homogeneous polarization models are difficult to reconcile with the expected strong tendency for domain formation<sup>8,9</sup>, but the effect of domains on negative capacitance has received little attention<sup>5,10–12</sup>. Here we report negative capacitance in a model system of multidomain ferroelectric–dielectric superlattices across a wide range of temperatures, in both the ferroelectric and paraelectric phases. Using a phenomenological model, we show that domain-wall motion not only gives rise to negative permittivity, but can also enhance, rather than limit, its temperature range. Our first-principles-based atomistic simulations provide detailed microscopic insight into the origin of this phenomenon, identifying the dominant contribution of near-interface layers and paving the way for its future exploitation.

Negative capacitance in ferroelectrics arises from the imperfect screening of the spontaneous polarization<sup>5,10,13,14</sup>. Imperfect screening is intrinsic to semiconductor–ferroelectric, and even metal–ferroelectric, interfaces because of their finite effective screening lengths<sup>15,16</sup>. Alternatively, imperfect screening can be engineered in a controlled manner by deliberately inserting a dielectric layer of relative permittivity  $\epsilon_d$  between the ferroelectric and the electrodes<sup>6</sup> (Fig. 1a). The physical separation of the ferroelectric bound charge from the metallic screening charges creates a depolarizing field inside the ferroelectric, destabilizing the polarization and lowering the ferroelectric transition temperature  $T_C$ . The effect of the dielectric layer can be understood by considering the free energy of the bilayer capacitor with the usual assumption of a uniform polarization  $P$  (see Methods). Below the bulk transition temperature  $T_0$ , the free energy of the ferroelectric layer develops a double-well with minima at finite values of  $P$ , but, when combined with the parabolic potential of the dielectric layer, the total energy has a minimum at  $P=0$  (Fig. 1b). The reciprocal dielectric constant of the system as a whole  $\epsilon^{-1}$ , given by the curvature of the total energy with respect to the polarization, is positive, as required for thermodynamic stability. However, because the non-polar state of the

ferroelectric layer corresponds to a maximum of its local energy, the local stiffness of the ferroelectric layer is negative; that is, polarizing the ferroelectric layer has a negative energy cost.



**Figure 1 | Phenomenological description of negative capacitance.** **a**, Sketch of the ferroelectric–dielectric bilayer capacitor with (right) and without (left) domains. Green, blue and grey layers correspond to the dielectric, ferroelectric and metallic components, respectively. **b**, The total (pink) and local free energies  $F$  of the ferroelectric (blue) and dielectric (green) layers. **c**, Temperature  $T$  dependence of the local dielectric stiffness of the ferroelectric layer calculated from phenomenological models with uniform homogeneous polarization (blue), and with inhomogeneous polarization with static, soft domain walls (red, dashed) and mobile, abrupt (thin) domain walls (red, solid). The dotted line marks the breakdown of the thin-wall Landau–Kittel model close to  $T_C$  (see Methods).  $P$ , polarization;  $T_C^h$  and  $T_C^{ih}$ , temperature of ferroelectric transitions to homogeneous and inhomogeneous states, respectively;  $T_0$ , bulk transition temperature;  $\epsilon_d$  and  $\epsilon_f$ , dielectric constants of the dielectric and ferroelectric layers, respectively.

<sup>1</sup>London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, 17–19 Gordon Street, London WC1H 0HA, UK. <sup>2</sup>Institut de Ciència de Materials de Barcelona (ICMAB-CSIC), Campus UAB, 08193 Bellaterra, Spain. <sup>3</sup>Department of Quantum Matter Physics, University of Geneva, CH-1211 Geneva, Switzerland. <sup>4</sup>Laboratory of Condensed Matter Physics, University of Picardie, Amiens 80000, France. <sup>5</sup>L. D. Landau Institute for Theoretical Physics, Moscow, Russia. <sup>6</sup>Materials Research and Technology Department, Luxembourg Institute of Science and Technology (LIST), 5 avenue des Hauts-Fourneaux, L-4362 Esch/Alzette, Luxembourg.

\*These authors contributed equally to this work.

With decreasing temperature, the ferroelectric double-well progressively deepens and would dominate the total energy below a temperature  $T_C^h$ , favouring a transition to a homogeneous ferroelectric state. The local dielectric stiffness of the ferroelectric layer would then increase and eventually become positive, as shown by the blue curve in Fig. 1c. This ‘homogeneous’ model has served as the basis for the interpretation of previous experimental studies of negative capacitance<sup>17–20</sup>. However, although attractively simple, it does not describe the true ground state of the system because, in general, the depolarization field that leads to the negative capacitance effect will also tend to favour a transition to an inhomogeneous, multidomain phase at a temperature  $T_C^i$  ( $T_C^i > T_C^h$ ), as demonstrated by numerous experiments (see, for example, ref. 8). This has consequences for the dielectric response and negative capacitance, as we show with the help of two phenomenological models.

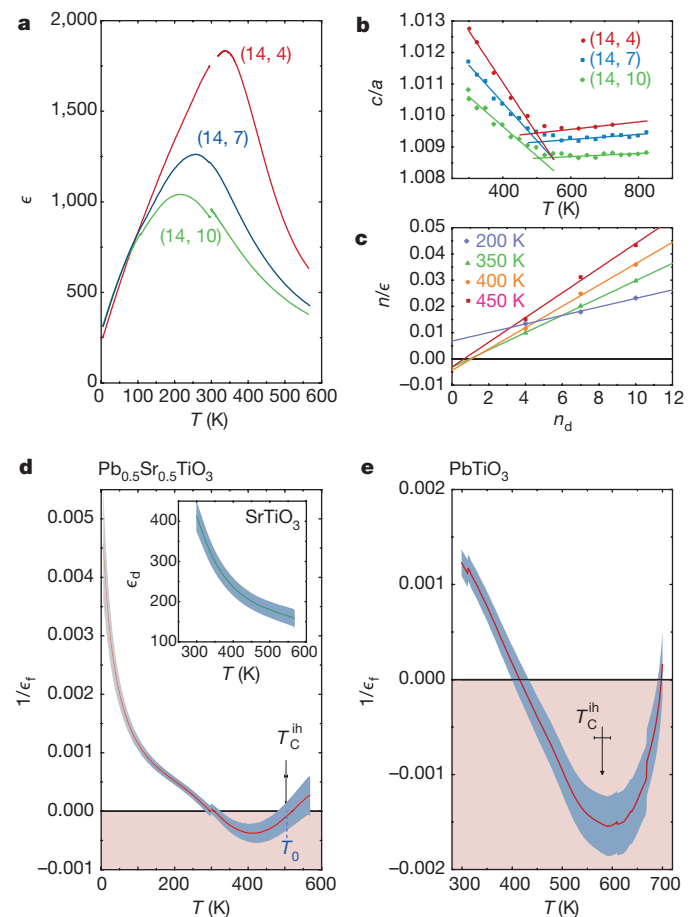
First, we use a Ginzburg–Landau approach (Methods) to obtain an analytic description of the phase transition into an inhomogeneous state with a gradual (soft) polarization profile, typical of ultrathin films<sup>21</sup>. This model allows us to obtain the lattice contribution to the dielectric response (that is, the response of a static domain structure; dashed curve in Fig. 1c). The appearance of the soft domain structure results in qualitative changes in the shape of the  $\epsilon^{-1}(T)$  curve, pushing its minimum below the actual transition temperature. However, the overall effect of a static domain structure is to reduce the temperature range of negative capacitance, as previously thought<sup>11</sup>.

Second, to investigate the contribution of domain-wall motion, we choose to work in the simpler Kittel approximation, which is valid for the abrupt (thin) domain walls typical of thicker films well below  $T_C$  (refs 5, 12, 22). The resulting dielectric response is shown by the solid red curve in Fig. 1c (for details of the calculation, see Methods). Remarkably, domain-wall motion contributes negatively to the overall dielectric stiffness<sup>5,10,12</sup>. Macroscopically, domain-wall displacements create a net polarization that leads to a depolarizing field, which dominates the total field in the ferroelectric, thus leading to negative capacitance. Microscopically, the domain-wall displacements redistribute the interfacial stray fields resulting in a negative net contribution to the free energy and thus to the local dielectric constant. Although the thin-wall Kittel model does not capture the subtleties of the soft domain structure of ultrathin films, it highlights the importance of the contribution of domain walls to extending the temperature window of negative capacitance.

To experimentally access the different temperature regimes of negative capacitance shown in Fig. 1c, we deposited several series of high-quality epitaxial superlattices consisting of  $n_f$  ferroelectric and  $n_d$  dielectric monolayers repeated  $N$  times, hereafter labelled  $(n_f, n_d)_N$ . For each superlattice series,  $n_f$  is fixed while  $n_d$  is varied from four to ten unit cells. SrTiO<sub>3</sub> crystals were used as substrates and epitaxial SrRuO<sub>3</sub> top and bottom electrodes were deposited *in situ* to enable dielectric impedance spectroscopy measurements. SrTiO<sub>3</sub> was also chosen as the dielectric component, and PbTiO<sub>3</sub> and quasi-random Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> alloys were used as the ferroelectric layers. The Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> composition was chosen for its low  $T_0$ , enabling us to investigate the full range of temperatures up to and above  $T_0$  without complications arising from leakage.

Such superlattices constitute a model system for the observation of negative capacitance, because they are mathematically equivalent to the bilayer systems investigated theoretically<sup>5,6,21</sup> and present a number of very convenient features—for example, the small layer thicknesses minimize the number of free carriers, ensuring appropriate electrostatic boundary conditions, while the highly ordered stripe domains are well-suited for X-ray diffraction (XRD) studies and theoretical modelling. By varying the thicknesses of the dielectric layers and the total number of bilayer repetitions, the permittivities of the individual layers can be extracted from measurements of the total capacitance of a series of samples.

The dielectric properties of three Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices with 14-unit-cell-thick Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> layers are summarized in Fig. 2a–d. All superlattices exhibit a broad maximum in the dielectric response that moves to lower temperature with increasing SrTiO<sub>3</sub> content (Fig. 2a). These maxima do not coincide with the phase-transition temperature  $T_C^h$  and instead arise from the qualitatively different temperature dependences of the permittivities of the SrTiO<sub>3</sub> and Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> layers. Using XRD, we obtain an estimate of  $T_C^h$  from the temperature evolution of in-plane and out-of-plane lattice parameters ( $a$  and  $c$ , respectively), as shown in Fig. 2b. Contrary to what is expected for a transition to a homogeneous ferroelectric state, the observed  $T_C^h$  is independent of the thickness of the SrTiO<sub>3</sub> layer because  $T_C^h$  is determined by the domain-wall density, which in turn depends only on the thickness of the ferroelectric layer. The regular domain structure with



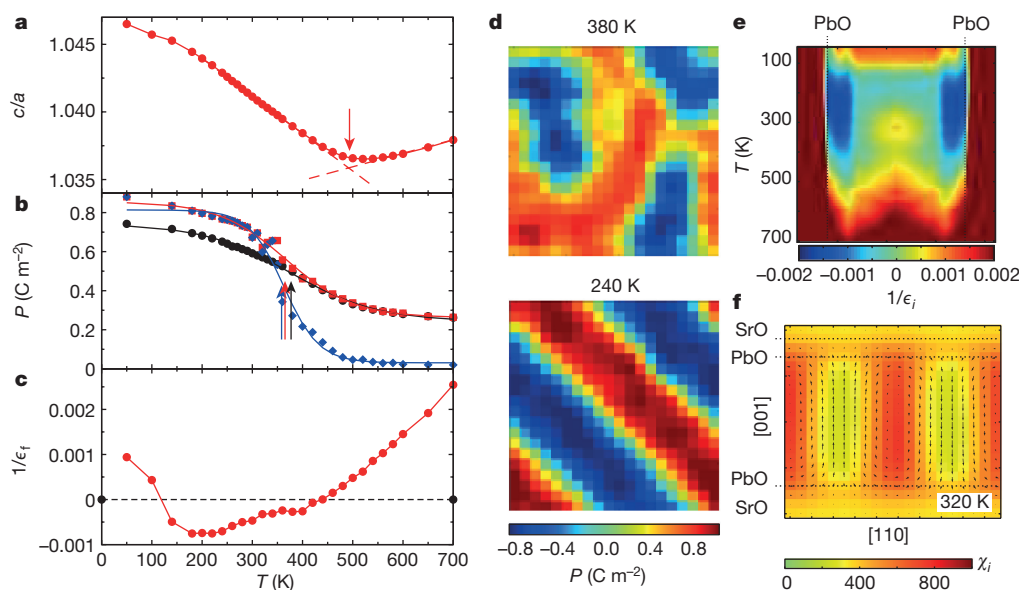
**Figure 2 | Temperature dependence of the dielectric permittivities of Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub>–SrTiO<sub>3</sub> and PbTiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices.** **a**, Total dielectric constant  $\epsilon$  of  $(14, n_d)$  Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices as a function of temperature  $T$ . Red,  $n_d = 4$ ; blue,  $n_d = 7$ ; green,  $n_d = 10$ . **b**, Sample tetragonalities  $c/a$  used to determine  $T_C$ . **c**, Linear fits to the series capacitor expression ( $n/\epsilon \approx n_d/\epsilon_d + n_f/\epsilon_f$ , where  $n = n_d + n_f$ ,  $\epsilon_d$  and  $\epsilon_f$  are the dielectric constants of the dielectric and ferroelectric layers, and  $n_d$  and  $n_f$  are the numbers of dielectric and ferroelectric monolayers per period of the superlattice) for a selection of temperatures. **d**, Reciprocal dielectric constant of the Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> layers in  $(14, n_d)$  superlattices calculated from the series capacitor model. Dashed line indicates our estimate of  $T_0$  obtained from the temperature dependence of  $c/a$  of a thin Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> film. Inset, dielectric constant of the SrTiO<sub>3</sub> layers. **e**, Reciprocal dielectric constant of the PbTiO<sub>3</sub> layers in  $(5, n_d)$  superlattices. In **d** and **e**, the arrows indicate  $T_C^h$  with associated error bars representing the spread of values between the samples in the series; grey shading indicates estimated uncertainties obtained from weighted-least-squares linear fitting with weights determined from inter-electrode variation at room temperature within each sample.

a periodicity of about 10–12 nm can be observed using XRD as peaks in the diffuse scattering around the superlattice Bragg reflections (Extended Data Fig. 1).

To separate the dielectric constants  $\epsilon_d$  and  $\epsilon_f$  of individual layers we apply the standard series capacitor expression, which for our superlattices is  $n/\epsilon \approx n_d/\epsilon_d + n_f/\epsilon_f$  (see Methods), where  $n = n_d + n_f$  and  $\epsilon$  is the overall dielectric constant of the superlattice, obtained directly from the measured capacitance. The linear relationship between  $n/\epsilon$  and  $n_d$  is well satisfied for  $100 \text{ K} \lesssim T < 570 \text{ K}$ , as illustrated in Fig. 2c for a few selected temperatures. The dielectric constant of the  $\text{SrTiO}_3$  layers can be obtained from the slopes of the plots in Fig. 2c. The resulting  $\epsilon_d(T)$  is presented in the inset of Fig. 2d and shows the typical decrease with temperature observed in  $\text{SrTiO}_3$  thin films and bulk crystals. The intercepts of the linear plots in Fig. 2c give the reciprocal dielectric constant of the  $\text{Pb}_{0.5}\text{Sr}_{0.5}\text{TiO}_3$  layers  $\epsilon_f^{-1}$ , which is plotted in Fig. 2d. At low temperature, deep in the ferroelectric regime,  $\epsilon_f^{-1}$  is positive. However, upon heating it slowly decreases, entering the negative capacitance regime around room temperature. It then reaches a minimum, and subsequently returns to positive values at high temperature in the paraelectric phase. The minimum in  $\epsilon_f^{-1}$  is observed well below the phase-transition temperature  $T_C^{\text{th}}$  (indicated with an arrow in Fig. 2d), contrary to what would be expected for a structure with homogeneous polarization. For this series of samples, the temperature regime  $T_C^{\text{th}} < T < T_0$  cannot be resolved because  $T_C^{\text{th}}$  is very close to  $T_0$  (measured independently to be around 500 K for  $\text{Pb}_{0.5}\text{Sr}_{0.5}\text{TiO}_3$  thin films of the same composition). To access this temperature regime, a set of  $(5, n_d)_N$  superlattices was fabricated with  $\text{Pb}_{0.5}\text{Sr}_{0.5}\text{TiO}_3$  replaced by  $\text{PbTiO}_3$ , which has a much higher  $T_0 \approx 1,200 \text{ K}$  when grown coherently on  $\text{SrTiO}_3$  (ref. 23). As shown in Fig. 2e, the negative capacitance regime can be clearly observed in the paraelectric phase above  $T_C^{\text{th}} \approx 580 \text{ K}$  in these samples. However, above  $T_C$  the dielectric stiffness increases much faster than expected, becoming positive far below  $T_0$ . This is probably due to the progressive increase in the thermally

activated conductivity of the ferroelectric layers, which destroys the electrostatic boundary conditions required for negative capacitance and leads to Maxwell–Wagner relaxations at high temperature (see Methods).

To gain further insight, we used first-principles-based effective models that permit the treatment of thermal effects. We used the potentials for  $\text{PbTiO}_3$  and  $\text{SrTiO}_3$  introduced in ref. 24 as the starting point to construct models for  $\text{PbTiO}_3$ – $\text{SrTiO}_3$  superlattices with an in-plane epitaxial constraint corresponding to a  $\text{SrTiO}_3(001)$  substrate (see Methods). As compared with experiment, our models feature relatively stiff  $\text{SrTiO}_3$  layers and relatively low ferroelectric transition temperatures; otherwise, they capture the behaviour of  $\text{PbTiO}_3$  layers stacked with dielectric layers in a qualitatively and semi-quantitatively correct way. For computational feasibility, we focused on a representative  $(8, 2)$  superlattice ( $10 \times 10 \times 10$  elemental perovskite cells in the periodically repeated simulation box) that presents the behaviour summarized in Fig. 3. As the superlattice is cooled from high temperature, the  $c/a$  ratio of the  $\text{PbTiO}_3$  layers (Fig. 3a) evidences an elastic transition, at about 490 K, to a state characterized by strongly fluctuating ferroelectric domains (380-K snapshot provided in Fig. 3d and Supplementary Video 1). This fluctuating phase could be indicative of temperature-induced domain melting, analogous to vortex lattice melting in high-temperature superconductors<sup>25</sup>. As we further cool the superlattice, we observe a ferroelectric transition at 370 K associated with the freezing of the domains into stable stripes. This change can be appreciated in Fig. 3b, in which we plot different measures of the local dipole order inside the  $\text{PbTiO}_3$  layer. As shown in the bottom panel of Fig. 3d, this low-temperature phase presents stripes along the  $[110]$  direction, with a domain thickness of about five unit cells and sharp walls. As shown in Fig. 3f and Extended Data Fig. 2, in the ground state, the dipoles form closure domains and almost do not penetrate into the stiff  $\text{SrTiO}_3$  layers. This corresponds to the vanishing of spontaneous polarization at the surface of a polydomain ferroelectric<sup>26</sup>. The domain



**Figure 3 | Results of Monte Carlo simulations of a first-principles-based model for the  $(8, 2)$  superlattice.** **a–c**, Temperature  $T$  dependence of the tetragonality  $c/a$  (**a**), local polarization  $P$  (**b**) and reciprocal dielectric constant  $1/\epsilon_f$  (**c**) of the  $\text{PbTiO}_3$  layer. Solid lines are guides to the eye. **a**, The dashed lines extrapolate the linear behaviour above and below the transition, and help us locate the elastic transition temperature. **b**, Supercell average of the absolute value of the local polarization (black circles), as well as the polarization at a particular cell within a domain considering its absolute (red squares) and bare (blue diamonds) values. The arrow in **a** marks the elastic transition and the onset of

fluctuating polar order around 490 K; those in **b** mark the ferroelectric freezing transition around 370 K as determined from the inflection points. The high-temperature tails of the black and red curves in **b** reveal the presence of incipient polar order. **d**, Snapshots of the local polarization  $P$  (out-of-plane component) within the middle of the  $\text{PbTiO}_3$  layer at 380 K (top) and 240 K (bottom). **e**, Temperature dependence of the local dielectric response  $1/\epsilon_f$  resolved along the stacking direction. **f**, Local susceptibility  $\chi_i$  map in the  $(\bar{1}10)$  plane at 320 K. The small arrows represent the projection in the  $(\bar{1}10)$  plane of the local electric dipoles as deduced from the equilibrium atomic configuration.

walls exhibit substantial Bloch character in the ground state; this is the result of a wall-confined polarization along  $[110]$  that appears at about 120 K and is analogous to the one recently predicted<sup>27</sup> for pure  $\text{PbTiO}_3$ .

We investigated the layer-resolved dielectric response of the superlattices. In essence (more details in Methods), we compute the local susceptibility of a region  $i$ ,  $\chi_i = \frac{1}{\epsilon_0} \frac{\partial \langle P_i \rangle}{\partial E_{\text{ext}}}$ , in which  $\epsilon_0$  is the vacuum permittivity,  $P_i$  is the local polarization and  $\langle \dots \rangle$  represents a thermal average that can be readily obtained by simulating our models under an applied electric field  $E_{\text{ext}}$ . As shown in Methods, the local dielectric constant can be expressed as  $\epsilon_i = \epsilon_{\text{tot}} / (\epsilon_{\text{tot}} - \chi_i)$ , where  $\epsilon_{\text{tot}}$  is the dielectric constant of the whole system. The results in Fig. 3c correspond to such a calculation for the  $\text{PbTiO}_3$  layers of the (8, 2) superlattice, and confirm the presence of a region of negative capacitance extending above and below the ferroelectric transition temperature.

However, it is not immediately obvious where the computed negative capacitance comes from. The local susceptibilities  $\chi_i$  are always positive in our calculations, confirming the expectation that an applied external field induces polarization changes that are parallel to it. By contrast, the local dielectric constant  $\epsilon_i$  measures a response to a local field that incorporates depolarizing fields, making its behaviour richer and its physical interpretation more challenging<sup>5</sup>. In particular,  $\epsilon_i$  will be negative if  $\chi_i > \epsilon_{\text{tot}}$ . Hence, the negative capacitance regions are those that are substantially more responsive than the system as a whole.

Our formalism allows us to map out the local response within the  $\text{PbTiO}_3$  layers and thus determine which regions are responsible for the negative capacitance behaviour. Figure 3e shows  $\epsilon_i^{-1}$  resolved along the superlattice-stacking direction and as a function of temperature. At high temperatures, the material behaves like a normal dielectric. Then, negative contributions to  $\epsilon_i^{-1}$  appear at about 550 K, well before any ordering occurs in the system; in that regime, the negative contribution is confined to the vicinity of the  $\text{PbTiO}_3$ – $\text{SrTiO}_3$  interface, and the response of the whole  $\text{PbTiO}_3$  layer continues to be positive. As temperature is further reduced, the negative capacitance region extends to the whole  $\text{PbTiO}_3$  layer. Eventually, at low temperature, the inner part of the  $\text{PbTiO}_3$  layer recovers a conventional dielectric behaviour that dominates the total response, even if our simulations reveal that a negative contribution from the interfaces still persists.

We can further map the susceptibility within the planes perpendicular to the stacking direction to quantify the contributions of domains and domain walls. Figure 3f shows representative results at 320 K. Predictably, we find that the susceptibility at the domain walls is much larger than at the domains. In other words, the field-induced polarization of the walls, which results in the growth or shrinkage of the domains, dominates the response. Further, the large response of the walls is much enhanced in the vicinity of the interfaces with the  $\text{SrTiO}_3$  layers. Hence, our simulations suggest that, below about 370 K, the domain-wall region near the interfaces dominates the negative capacitance of the  $\text{PbTiO}_3$  layers.

There are important differences between our simulated and experimental superlattices that complicate a detailed comparison (see Methods for more details). Nevertheless, our basic result—that the  $\text{PbTiO}_3$  layers have a negative dielectric constant in a temperature region extending above and below  $T_C$ —is confirmed by our simulations. Further, we also ran simulations of various (8,  $n_d$ ) superlattices to mimic our experimental approach for calculating the response of the  $\text{PbTiO}_3$  layer; the results shown in Extended Data Fig. 3 are similar to those of Fig. 3c, thus validating our strategy to measure  $\epsilon_f$ .

Finally, the depolarization effects in ferroelectric–dielectric superlattices are completely analogous to those at interfaces between a ferroelectric and a metal or a semiconductor. We have found that  $\text{PbTiO}_3$ – $\text{SrRuO}_3$  superlattices, for example, exhibit very similar domain structures as  $\text{PbTiO}_3$ – $\text{SrTiO}_3$ . These structures are induced by the imperfect screening at the  $\text{SrRuO}_3$ – $\text{PbTiO}_3$  interfaces, which produces a depolarizing field equivalent to that induced by a 7-unit-cell-thick  $\text{SrTiO}_3$  layer<sup>28,29</sup>. It is therefore reasonable to expect a negative

capacitance effect of the same order of magnitude in a transistor-like structure composed of a  $\text{PbTiO}_3$  gate dielectric and an ultrathin conducting  $\text{SrRuO}_3$  channel, where applying a gate voltage  $V_g$  will lead to an enhancement of the surface potential  $\phi_s$  at the  $\text{PbTiO}_3$ – $\text{SrRuO}_3$  interface. With a  $\text{PbTiO}_3$ – $\text{SrRuO}_3$  interface capacitance  $C_i$  of about  $0.6 \text{ F m}^{-2}$  (ref. 14) and a ferroelectric capacitance  $C_f$  equivalent to that of one of our  $\text{PbTiO}_3$  layers, one can obtain voltage amplification factors  $\frac{\partial \phi_s}{\partial V_g} = \frac{C_f}{C_i + C_f}$  as large as about two at temperatures at which  $\epsilon_f^{-1}$  is most negative. For the more practical interface with a conventional semiconductor, the expected amplification is more modest (for example,  $\frac{\partial \phi_s}{\partial V_g} \approx 1.03$  for  $C_i \approx 0.1 \text{ F m}^{-2}$ , ref. 11), but is still enhanced compared to conventional gate dielectrics for which the corresponding value is less than unity. Such enhancements are especially encouraging in light of recent progress in the integration of ferroelectric oxides directly on conventional semiconductors<sup>30</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 17 July 2015; accepted 8 March 2016.**

**Published online 13 June 2016.**

1. Scott, J. F. & Paz de Araujo, C. A. Ferroelectric memories. *Science* **246**, 1400–1405 (1989).
2. Naumov, I. I., Bellaiche, L. & Fu, H. Unusual phase transitions in ferroelectric nanodisks and nanorods. *Nature* **432**, 737–740 (2004).
3. Garcia, V. *et al.* Giant tunnel electroresistance for non-destructive readout of ferroelectric states. *Nature* **460**, 81–84 (2009).
4. Kim, D. J. *et al.* Ferroelectric tunnel memristor. *Nano Lett.* **12**, 5697–5702 (2012).
5. Bratkovsky, A. M. & Levanyuk, A. P. Very large dielectric response of thin ferroelectric films with the dead layers. *Phys. Rev. B* **63**, 132103 (2001).
6. Salahuddin, S. & Datta, S. Use of negative capacitance to provide voltage amplification for low power nanoscale devices. *Nano Lett.* **8**, 405–410 (2008).
7. Krowne, C. M., Kirchoefer, S. W., Chang, W., Pond, J. M. & Alldredge, L. M. B. Examination of the possibility of negative capacitance using ferroelectric materials in solid state electronic devices. *Nano Lett.* **11**, 988–992 (2011).
8. Fong, D. D. *et al.* Ferroelectricity in ultrathin perovskite films. *Science* **304**, 1650–1653 (2004).
9. Catalan, G., Jiménez, D. & Gruverman, A. Ferroelectrics: Negative capacitance detected. *Nat. Mater.* **14**, 137–139 (2015).
10. Bratkovsky, A. M. & Levanyuk, A. P. Depolarizing field and “real” hysteresis loops in nanometer-scale ferroelectric films. *Appl. Phys. Lett.* **89**, 253108 (2006).
11. Cano, A. & Jiménez, D. Multidomain ferroelectricity as a limiting factor for voltage amplification in ferroelectric field-effect transistors. *Appl. Phys. Lett.* **97**, 133509 (2010).
12. Luk'yanchuk, I., Pakhomov, A., Sené, A., Sidorkin, A. & Vinokur, V. Terahertz electrodynamics of  $180^\circ$  domain walls in thin ferroelectric films. Preprint at <http://arxiv.org/abs/1410.3124> (2014).
13. Ponomareva, I., Bellaiche, L. & Resta, R. Dielectric anomalies in ferroelectric nanostructures. *Phys. Rev. Lett.* **99**, 227601 (2007).
14. Stengel, M., Vanderbilt, D. & Spaldin, N. A. Enhancement of ferroelectricity at metal–oxide interfaces. *Nat. Mater.* **8**, 392–397 (2009).
15. Mehta, R. R., Silverman, B. D. & Jacobs, J. T. Depolarization fields in thin ferroelectric films. *J. Appl. Phys.* **44**, 3379–3385 (1973).
16. Junquera, J. & Ghosez, P. Critical thickness for ferroelectricity in perovskite ultrathin films. *Nature* **422**, 506–509 (2003).
17. Khan, A. I. *et al.* Experimental evidence of ferroelectric negative capacitance in nanoscale heterostructures. *Appl. Phys. Lett.* **99**, 113501 (2011).
18. Appleby, D. J. R. *et al.* Experimental observation of negative capacitance in ferroelectrics at room temperature. *Nano Lett.* **14**, 3864–3868 (2014).
19. Gao, W. *et al.* Room-temperature negative capacitance in a ferroelectric–dielectric superlattice heterostructure. *Nano Lett.* **14**, 5814–5819 (2014).
20. Khan, A. I. *et al.* Negative capacitance in a ferroelectric capacitor. *Nat. Mater.* **14**, 182–186 (2015).
21. Luk'yanchuk, I. A., Lahoche, L. & Sené, A. Universal properties of ferroelectric domains. *Phys. Rev. Lett.* **102**, 147601 (2009).
22. Kopal, A., Mokry, P., Fousek, J. & Bahnik, T. Displacements of  $180^\circ$  domain walls in electroded ferroelectric single crystals: the effect of surface layers on restoring force. *Ferroelectrics* **223**, 127–134 (1999).
23. Dawber, M. *et al.* Tailoring the properties of artificially layered ferroelectric superlattices. *Adv. Mater.* **19**, 4153–4159 (2007).
24. Wojdeł, J. C., Hermet, P., Ljungberg, M. P., Ghosez, P. & Íñiguez, J. First-principles model potentials for lattice-dynamical studies: general methodology and example of application to ferroic perovskite oxides. *J. Phys. Condens. Matter* **25**, 305401 (2013).

25. Blatter, G., Feigel'man, M. V., Geshkenbein, V. B., Larkin, A. I. & Vinokur, V. M. Vortices in high-temperature superconductors. *Rev. Mod. Phys.* **66**, 1125–1388 (1994).
26. De Guerville, F., Luk'yanchuk, I., Lahoche, L. & El Marssi, M. Modeling of ferroelectric domains in thin films and superlattices. *Mater. Sci. Eng. B* **120**, 16–20 (2005).
27. Wojdel, J. C. & Íñiguez, J. Ferroelectric transitions at ferroelectric domain walls found from first principles. *Phys. Rev. Lett.* **112**, 247603 (2014).
28. Lichtensteiger, C., Fernandez-Pena, S., Weymann, C., Zubko, P. & Triscone, J.-M. Tuning of the depolarization field and nanodomain structure in ferroelectric thin films. *Nano Lett.* **14**, 4205–4211 (2014).
29. Aguado-Puente, P. & Junquera, J. Ferromagneticlike closure domains in ferroelectric ultrathin films: first-principles simulations. *Phys. Rev. Lett.* **100**, 177601 (2008).
30. Warusawithana, M. P. *et al.* A ferroelectric oxide made directly on silicon. *Science* **324**, 367–370 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We acknowledge financial support from the EPSRC (Grant No. EP/M007073/1; P.Z. and M.H.), the A. G. Leventis Foundation (M.H.); FNR Luxembourg (Grant No. FNR/P12/4853155/Kreisel; J.I.), MINECO-Spain (Grant No. MAT2013-40581-P; J.I. and J.C.W.), the Swiss National Science Foundation Division II (J.-M.T. and S.F.-P.), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC (Grant No. 319286 (Q-MAC); J.-M.T. and S.F.-P.), and the EU-FP7-ITN project NOTEDEV (Grant No. 607521; I.L.).

**Author Contributions** P.Z., M.H., S.F.-P. and J.-M.T. performed and analysed the experiments. A.S. and I.L. developed the phenomenological theory. J.C.W. and J.I. developed the atomistic models and performed the simulations.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.Z. ([p.zubko@ucl.ac.uk](mailto:p.zubko@ucl.ac.uk)) or J.I. ([jorge.iniguez@list.lu](mailto:jorge.iniguez@list.lu)).

## METHODS

**Landau theory for monodomain bilayers and superlattices.** To derive the expected temperature dependence of the dielectric function of a ferroelectric–dielectric bilayer or superlattice undergoing a phase transition to a homogenous (monodomain) state, we consider the free energy of the bilayer capacitor under short-circuit boundary conditions (or equivalently, one period of a superlattice) of the form

$$F = l_f \left( \frac{\alpha_f}{2} P^2 + \frac{\beta_f}{4} P^4 + \frac{\epsilon_0}{2} E_f^2 \right) + l_d \frac{\epsilon_0 \epsilon_d}{2} E_d^2 \quad (1)$$

The first term represents the energy density, per unit area, of a ferroelectric material with a second-order phase transition at a temperature  $T_0$  as determined by the coefficient of the  $P^2$  term,  $\alpha_f = (T - T_0)/(C\epsilon_0)$ . The second term describes the energy penalty for polarizing the dielectric layer with dielectric constant  $\epsilon_d$ . Here  $E_f$  and  $E_d$  are the electric fields appearing in the ferroelectric and dielectric layers, respectively, when the spontaneous polarization  $P$  develops,  $C$  is the Curie constant,  $\beta_f$  is the coefficient of the  $P^4$  term and  $l_{d,f}$  are the thicknesses of the dielectric and ferroelectric layers respectively. Taking into account the electrostatic boundary conditions at the ferroelectric–dielectric interface,  $\epsilon_d \epsilon_0 E_d = \epsilon_0 E_f + P$ , and the short-circuit condition for the whole system,  $l_d E_d + l_f E_f = 0$ , the functional in equation (1) can be rewritten in terms of only  $P$  with a renormalized overall  $P^2$  coefficient and the corresponding lowering of the transition temperature. The transition to a homogeneous ferroelectric state is thus predicted to occur at a temperature

$$T_C^h = T_0 - C \left( 1 + \frac{l_f}{l_d} \epsilon_d \right)^{-1}$$

In particular, when  $\epsilon_d \gg 1$  and  $l_d$  is of the order of  $l_f$ , equation (1) reduces to

$$F \approx l_f \left( \frac{\alpha_f}{2} P^2 + \frac{\beta_f}{4} P^4 \right) + l_d \left( \frac{1}{2\epsilon_0 \epsilon_d} P^2 \right)$$

which describes the energy of a homogeneously polarized bilayer with equal polarizations in both layers<sup>17,23</sup>.  $T_C^h$  then simplifies to

$$T_C^h \approx T_0 - \frac{l_d}{l_f \epsilon_d} C$$

The overall electric susceptibility  $\chi$  of such a system is given by

$$\frac{l}{\chi} = \epsilon_0 \frac{\partial^2 F}{\partial P^2} = \frac{l_d}{\chi_d} + \frac{l_f}{\chi_f}$$

where  $l = l_f + l_d$ , and  $\chi_d \approx \epsilon_d$  and  $\chi_f = (\alpha_f + 3\beta_f P^2)^{-1} \epsilon_0^{-1}$  are the electric susceptibilities of the dielectric and ferroelectric layers, respectively. It has the familiar form of the series capacitance formula,  $1/C_{\text{tot}} = 1/C_d + 1/C_f$ . For high permittivity materials such as those considered in this work,  $\chi = \epsilon$  is a very good approximation. The temperature dependence of the contribution to the reciprocal dielectric constant from the ferroelectric layer is shown in Fig. 1c (blue curve). Although the total permittivity (not shown) exhibits the typical divergence ( $\epsilon^{-1} = 0$ ) at  $T_C^h$  and is always positive, as required for thermodynamic stability, the dielectric stiffness of the ferroelectric component decreases linearly with temperature upon cooling and acquires negative values below  $T_0$ . At  $T_C^h$ , the spontaneous polarization appears and the  $3\beta_f P^2$  term eventually restores  $\epsilon_f^{-1}$  to positive values at lower temperatures. To obtain the blue curve in Fig. 1c, we modelled a 30-nm-thick PbTiO<sub>3</sub> film in series with a 10-nm-thick SrTiO<sub>3</sub> layer using the following parameters:  $T_0 = 1,244$  K (strain-renormalized),  $C = 4.1 \times 10^5$  K and  $\epsilon_d = 300$ , giving  $T_C^h = 788$  K.

**Landau–Kittel model of domain-wall contribution to permittivity.** For an isolated ferroelectric slab of thickness  $l_f$  in zero applied field, the up- and down-oriented 180° domains are of equal width  $w$ , given by the Landau–Kittel square-root dependence<sup>31,32</sup>. For high- $\epsilon$  ferroelectrics

$$w \approx \left( \frac{\epsilon_{\perp}}{\epsilon_{\parallel}} \right)^{\frac{1}{4}} \sqrt{\zeta \times \lambda \times 2\zeta \times l_f}$$

where  $\epsilon_{\parallel}$  and  $\epsilon_{\perp}$  are the ‘bulk’ lattice dielectric constants parallel and perpendicular to the polarization, respectively,  $\xi$  is the coherence length,  $\lambda = 1 + \epsilon_d/(\epsilon_{\parallel}\epsilon_{\perp})^{1/2}$  and  $\zeta \approx 3.53$  (refs 5, 12, 26, 33, 34). This equation also holds for ferroelectric films with ‘dead layers’ and ferroelectric–dielectric superlattices, provided that the dielectric layers are thick enough compared to the domain width to allow the interfacial stray fields to decay sufficiently. Upon application of a field, the ferroelectric layer develops a net polarization due to the dielectric response of the lattice, described by  $\epsilon_{\parallel}$ , and to the motion of domain walls. To calculate the domain-wall contribution,

one must find the field-induced changes to the stray depolarizing fields, as has been done in refs 5, 12, 22. The resulting effective dielectric constant of the ferroelectric can be expressed as<sup>12</sup>

$$\epsilon_f = \epsilon_{\parallel} - \frac{\pi}{4\ln(2)} \sqrt{\frac{\epsilon_{\perp}}{\epsilon_{\parallel}}} \frac{l_f}{w} \lambda \epsilon_{\parallel}$$

where the first term is the lattice response and the second term is the negative contribution from domain-wall motion. Within the limits of the validity of the Landau–Kittel theory,  $l_f/w$  is large and therefore the second term is dominant. This term originates from the field-induced changes in the inhomogeneous electric-field distribution at the interface between the ferroelectric and the dielectric layers, consistent with the findings of our atomistic calculations.

The temperature dependence of  $w$  and  $\epsilon_f$  can be estimated<sup>35</sup> using the standard critical Ginzburg–Landau expansions near  $T_0$

$$\xi(T) = \frac{\xi_0}{(1 - T/T_0)^{1/2}}; \quad \epsilon_{\parallel}(T) = \frac{1}{2} \frac{\kappa_{\parallel}}{1 - T/T_0} \quad (2)$$

where  $\kappa_{\parallel}$  is related to the Curie constant  $C$  via  $\kappa_{\parallel} = C/T_0$  and  $\xi_0$  is the atomic-scale coherence length at  $T = 0$ . Assuming that  $\epsilon_{\perp}$  is temperature independent, the domain width  $w$  is almost temperature independent<sup>21,26</sup>, whereas the approximate temperature dependence of  $\epsilon_f$  is sketched in Fig. 1c. The solid red curve in Fig. 1c was calculated for a 30-nm-thick film with the following parameters (corresponding roughly to those of strained PbTiO<sub>3</sub>):  $T_0 = 1,244$  K,  $C = 4.1 \times 10^5$  K,  $\epsilon_{\perp} = 120$  and  $2\xi_0 = 1$  nm.

**Ginzburg–Landau theory of polydomain bilayers and superlattices.** The critical temperature of transition to the inhomogeneous striped domain state can be calculated within Ginzburg–Landau theory<sup>26,33</sup>

$$T_C^{\text{th}} = (1 - \pi\tau) T_0$$

where  $\tau = (C/T_0\epsilon_{\perp})^{1/2} \times 2\xi_0/l_f$ . For a 30-nm-thick PbTiO<sub>3</sub> film, we obtain  $T_C^{\text{th}} \approx 1,030$  K. A similar expression (up to a numerical factor) can be obtained on the qualitative level by noting that, at  $T_C^{\text{th}}$ , the domain width  $w$  becomes comparable with the domain-wall thickness  $2\xi(T)$ .

Close to  $T_C^{\text{th}}$ , the Landau–Kittel thin-wall approximation breaks down as the domain profile becomes soft (we represent this region by the dotted line in Fig. 1c). The theory for mobile domain walls in this regime is challenging, but the lattice part of the response of the polydomain structure can be calculated analytically. This would correspond to a situation in which domain-wall motion is impeded, for instance, by pinning of the domain walls. Using Ginzburg–Landau theory, this contribution can be expressed as<sup>35</sup>

$$\epsilon_f(T) = \frac{2\epsilon_{\parallel}(T)}{3\langle P^2 \rangle / P_B^2 - 1}$$

which is a generalization of equation (2). Here,  $P_B = P_0(1 - T/T_0)^{1/2}$  is the normalized temperature-dependent polarization of the bulk short-circuited sample and  $\langle P^2 \rangle = \langle P^2(x, y) \rangle$  is the spatial average of the temperature-dependent polarization profile of the domain state with critical temperature  $T_C^{\text{th}}$ . The factor  $\langle P^2 \rangle$  can be calculated over a wide temperature interval that includes both soft and abrupt (thin) domain profiles using the universal expression for  $P(x, y)$  in terms of elliptic sn functions, as given in equation (7) of ref. 21. After spatially averaging we obtain

$$\epsilon_{f,\text{GL}}^{-1}(T) = \frac{T_0}{C} \left( \frac{T}{T_0} - 1 \right) \left[ 1 - 3F \left( 1 - \frac{T}{T_C^{\text{th}}} \right) \right], \quad T < T_C^{\text{th}}$$

where  $F(x) = [K(m(x)) - E(m(x))]^2 / m^2 \times \tanh(0.357\tau x)$ ,  $K(m)$  and  $E(m)$  are the complete elliptic integrals of the first and second kind, respectively, with  $m(x) \approx \tanh(0.27\tau x)$  and  $\tau = (C/T_0\epsilon_{\perp})^{1/2} \times 2\xi_0/l_f$ . Note that  $F(0) = 0$  and that the above expression matches the relative permittivity of the paraelectric state  $\epsilon_f(T) = C/(T - T_0)$  at  $T = T_C^{\text{th}}$ . The temperature dependence of  $\epsilon_{f,\text{GL}}^{-1}$  is shown by the dashed red line in Fig. 1c.

**Sample preparation.** Superlattices were deposited on monocrystalline SrTiO<sub>3</sub>(100) substrates using off-axis radiofrequency magnetron sputtering. PbTiO<sub>3</sub> and SrTiO<sub>3</sub> were deposited at a substrate temperature of 520 °C in an O<sub>2</sub>/Ar mixture of ratio 5/7 and total pressure of 180 mTorr. For SrRuO<sub>3</sub> layers, acting as top and bottom electrodes, the corresponding parameters were 635 °C, 1/20 and 100 mTorr. Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> layers were deposited by sequential sputtering of sub-monolayer amounts of SrTiO<sub>3</sub> and PbTiO<sub>3</sub>. The Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices were asymmetrically terminated with bottom SrRuO<sub>3</sub>–Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> and top SrTiO<sub>3</sub>–SrRuO<sub>3</sub> interfaces. By contrast, the PbTiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices were

symmetrically terminated with both metal–insulator interfaces being between SrTiO<sub>3</sub> and SrRuO<sub>3</sub>; the thickness of interfacial SrTiO<sub>3</sub> layers was chosen to be half of those in the superlattice interior to maintain a constant overall composition. For each series of superlattices, the thickness of the ferroelectric layers was fixed (14 unit cells for Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub> and 5 unit cells for PbTiO<sub>3</sub>), whereas the thickness of the SrTiO<sub>3</sub> layers was varied from 4 unit cells to 10 unit cells. The number of repetitions  $N$  was chosen to maintain the total superlattice thickness as close as possible to 100 nm for PbTiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices and 200 nm for Pb<sub>0.5</sub>Sr<sub>0.5</sub>TiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices. To extract the interface capacitance contribution, a series of (5, 8)<sub>N</sub> PbTiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices with  $N = 10, 19$  and 30 was used.

The top SrRuO<sub>3</sub> layers were patterned using ultraviolet photolithography and etched using an Ar ion beam to form a series of 240 μm × 240 μm capacitors. Structural characterization was performed using a PANalytical X'Pert PRO diffractometer equipped with a triple axis detector and an Anton Paar domed heating stage. Dielectric impedance spectroscopy in the 100-Hz to 2-MHz frequency range was performed using an Agilent E4980A Precision LCR meter in a tube furnace with a custom-made sample holder under continuous O<sub>2</sub> flow at atmospheric pressure.

**Structural analysis.** Specular  $\theta$ – $2\theta$  scans were used to determine the periodicity of the superlattice (Extended Data Fig. 1a), whereas rocking curves were used to confirm the presence of domains and determine their periodicity (Extended Data Fig. 1b, c). Temperature evolution of the lattice parameters was obtained from  $\theta$ – $2\theta$  scans and used to determine the phase-transition temperatures, taken to be the crossing point of linear fits to the high and low temperature data (see Fig. 2b). **Calculation of the permittivities of the individual layers.** The total measured capacitance of the sample  $C_{\text{tot}}$  has contributions from the superlattice  $C_{\text{SL}}$  and the two metal–dielectric interfaces  $C_i$

$$\frac{1}{C_{\text{tot}}} = \frac{1}{C_{\text{SL}}} + \frac{2}{C_i} = \frac{1}{C_f} + \frac{1}{C_d} + \frac{2}{C_i}$$

where  $C_f$  and  $C_d$  are the total (series) capacitances of all the ferroelectric (PbTiO<sub>3</sub>) and dielectric (SrTiO<sub>3</sub>) layers, respectively. For an ( $n_f, n_d$ )<sub>N</sub> superlattice

$$\frac{d}{\epsilon} = \frac{d_d}{\epsilon_d} + \frac{d_f}{\epsilon_f} + \frac{2}{\tilde{C}_i}$$

$$\frac{N(n_f + n_d) \tilde{c}}{\epsilon} = \frac{Nn_d c_d}{\epsilon_d} + \frac{Nn_f c_f}{\epsilon_f} + \frac{2}{\tilde{C}_i}$$

where  $\tilde{C}_i = C_i / (\epsilon_0 A)$ ,  $A$  is the sample area,  $d_{d,f}$  are the total thicknesses of the dielectric and ferroelectric components, respectively,  $d = d_d + d_f$  and  $c_{d,f}$  are the lattice constants of the dielectric and ferroelectric layers, respectively. Because  $c_f \approx c_d \approx \tilde{c} \equiv (n_d c_d + n_f c_f) / (n_d + n_f)$

$$\frac{n_d + n_f}{\epsilon} \equiv \frac{n}{\epsilon} \approx \frac{n_d}{\epsilon_d} + \frac{n_f}{\epsilon_f} + \frac{2}{N\tilde{c}\tilde{C}_i}$$

For a series of superlattices with a fixed  $n_f$  and  $n_d$ , but varying  $N$ , the interfacial contribution  $1/(\tilde{c}\tilde{C}_i)$  can be obtained from the slope of a plot of  $n/\epsilon$  versus  $1/N$ . Once the temperature dependence of the interfacial capacitance is known, the permittivities of the individual SrTiO<sub>3</sub> and PbTiO<sub>3</sub> layers can be obtained, using a series of samples with fixed  $n_f$  and varying  $n_d$ , from the slope and intercept of the plot of  $n/\epsilon - 2/(N\tilde{c}\tilde{C}_i)$  versus  $n_d$ .

This analysis relies on the assumption that the layer permittivities do not change as the thicknesses of the individual layers are varied within each superlattice series. It is therefore crucial that the thickness of the ferroelectric layers is held fixed, because it determines the periodicity of the ferroelectric domain structure and thus the ferroelectric transition temperature and the domain-wall contribution to the measured dielectric constant. All superlattices within a series must also be in the same regime of electrostatic coupling, which places a lower limit on the thickness of the SrTiO<sub>3</sub> layers of around 3–4 unit cells<sup>36</sup>.

To quantify the interface contribution  $C_i$  for the PbTiO<sub>3</sub>-based superlattices, a series of symmetrically terminated samples with a fixed period (5, 8)<sub>N</sub>, but varying number of repetitions  $N$ , was fabricated. The interface capacitance was extracted from the intercept of the plot of  $n/\epsilon$  versus  $1/N$ , as discussed above, and is shown as a function of temperature in Extended Data Fig. 4. At room temperature,  $C_i$  is around 1,000 fF μm<sup>−2</sup>, which is in excellent agreement with previous experimental work<sup>37</sup> and compares quite well with the density functional theory prediction of 615 fF μm<sup>−2</sup> (at 0 K) for the same interface<sup>38</sup>. The weak dependence of  $C_i$  on temperature is also consistent with previous reports<sup>37</sup>. Quantifying the interfacial contribution independently in this way allows us to extract the temperature range of the negative capacitance regime more reliably. As illustrated in Extended Data Fig. 4, the interfacial contribution does not change the qualitative behaviour of the extracted PbTiO<sub>3</sub> dielectric constant and makes only a small (within error bars)

difference to the extracted PbTiO<sub>3</sub> stiffness. It is thus reasonable to neglect this correction, as was done in Fig. 2.

**Impedance analysis.** The observation of negative capacitance relies on the electrostatic interactions between the ferroelectric and dielectric layers, which in turn require both materials to be sufficiently insulating to avoid the screening of the spontaneous polarization. To identify the origin of dielectric losses and to quantify the conductivity of our samples, we measured complex impedance spectra over a wide range of frequencies from 100 Hz to 2 MHz and performed equivalent-circuit modelling. We present the complex impedance  $Z(\omega) = Z' + iZ''$  (in which  $\omega$  is the angular frequency,  $Z' = \text{Re}(Z)$  and  $Z'' = \text{Im}(Z)$ ) data in the complex capacitance representation  $C(\omega) = C' + iC'' \equiv 1/[i\omega Z(\omega)]$  (with  $C' = \text{Re}(C)$  and  $C'' = \text{Im}(C)$ ) as is common for capacitive systems.

Each PbTiO<sub>3</sub> and SrTiO<sub>3</sub> layer in the superlattice, as well as the two metal–dielectric interfaces, can be considered as a parallel R–C element, with a capacitance  $C_j$  and a resistance  $R_j$  due to the finite conductivity of the layer. The superlattice is then modelled by connecting these R–C elements in series, as shown in the inset of Extended Data Fig. 5. An additional series resistance  $R_s$  (typically a few hundred ohms) accounts for the contact resistances and other sources of resistance in the external circuit.

At low temperature, the conductivities of the PbTiO<sub>3</sub> and SrTiO<sub>3</sub> layers are negligible and the whole system behaves as a single capacitance  $C = (\sum_j C_j^{-1})^{-1}$ . The measured  $C'$  is frequency independent except for the high-frequency roll-off due to the parasitic series resistance  $R_s$ . As shown in Extended Data Fig. 5 for a (5, 8)<sub>30</sub> PbTiO<sub>3</sub>–SrTiO<sub>3</sub> superlattice, even at 500 K, the data can be well modelled by a single capacitor in series with  $R_s$ ; the parallel resistance is too high to be determined from the fit (that is, well above 10<sup>8</sup> Ω). At higher temperatures, the superlattice conductivity increases resulting in an increase in the dielectric loss  $C''$  at low frequencies. The 600-K data are modelled with one parallel R–C element in series with  $R_s$ . Despite the high temperature and large electrode area (240 μm × 240 μm), the total sample resistance is still 2 MΩ. However, at 700 K, the total sample resistance drops to 8.8 kΩ. In addition, some layers become substantially more conducting than others, giving rise to Maxwell–Wagner relaxations<sup>39</sup>, which can be observed as steps and plateaus in  $C'(\omega)$ . The behaviour can be qualitatively captured by dividing the system into two blocks with different resistances, each modelled as a parallel R–C element. However, to reproduce the more gradual frequency dispersion, more R–C elements are needed (in this case three were sufficient). At these temperatures, the samples are too conducting to maintain the electrostatic conditions necessary for negative capacitance. The sample resistances for all data shown in Fig. 2 were higher than 1 MΩ.

**Atomistic simulations of PbTiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices.** To construct the first-principles models for the PbTiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices, we took advantage of previously introduced<sup>24</sup> potentials for the bulk compounds, which give a qualitatively correct description of the lattice-dynamical properties and structural phase transitions of both materials. Then, we treated the interface between PbTiO<sub>3</sub> and SrTiO<sub>3</sub> in an approximate way, relying on the following observations. (1) The inter-atomic force constants in perovskite oxides such as PbTiO<sub>3</sub> and SrTiO<sub>3</sub> have been shown to depend strongly on the identity of the involved chemical species and weakly on the chemical environment<sup>40</sup>. (Hence, for example, the interactions between Ti and O are very similar in both PbTiO<sub>3</sub> and SrTiO<sub>3</sub>.) (2) Except in the limit of very-short-period superlattices, the main effects of the stacking are purely electrostatic and largely independent of the details of the interactions at the interfaces. (3) The main purely interfacial effects leading, for example, to the occurrence of new orders (such as those discussed in ref. 41) are related to the symmetry breaking, which permits new couplings that are forbidden by symmetry in the bulk case. Such qualitative symmetry-breaking effects are trivially captured by our potentials, even if the actual values of the interactions are approximate. (Similar approaches to treat ferroelectric superlattices and junctions can be found in the literature, ref. 42 being a representative case.)

As a result of these approximations, we were able to construct our superlattice potentials by using the models for bulk PbTiO<sub>3</sub> and SrTiO<sub>3</sub> to describe the interactions within the layers, assuming a simple numerical average for the interactions of the ion pairs touching or crossing the interface. For example, Ti–O interactions in a TiO<sub>2</sub> interface plane are computed as the average of the analogous Ti–O interactions in PbTiO<sub>3</sub> and SrTiO<sub>3</sub>. New interactions, such as those involving Pb and Sr neighbours across the interface, are chosen so that the acoustic sum rules are respected; in practice, their values are close to an average between the analogous Sr–Sr and Pb–Pb pairs. Finally, the long-range dipole–dipole interactions are governed by a bare electronic dielectric constant  $\epsilon_\infty$  that is taken as a weighted average of the first-principles results for bulk PbTiO<sub>3</sub> (8.5 $\epsilon_0$ ) and SrTiO<sub>3</sub> (6.2 $\epsilon_0$ ), with weights reflecting the composition of the superlattice.

The parameters of our models for bulk PbTiO<sub>3</sub> and SrTiO<sub>3</sub> were computed from first principles as described in ref. 24. To model our PbTiO<sub>3</sub>–SrTiO<sub>3</sub> superlattices, we adjusted our models in the following ways. (1) We softened the model for

bulk SrTiO<sub>3</sub> so that it has a dielectric permittivity  $\epsilon_{33}$  of about  $300\epsilon_0$  at room temperature. We checked *a posteriori* that the SrTiO<sub>3</sub> layers in the superlattices are not as soft, which is probably a consequence of the modified electrostatic interactions ( $\epsilon_\infty$ ) assumed, as described above. (2) We imposed an epitaxial constraint corresponding to having a SrTiO<sub>3</sub> (001)-oriented substrate; that is, we assume in-plane lattice constants  $a = b = 3.901 \text{ \AA}$ , forming an angle  $\gamma = 90^\circ$ . (3) We tweaked the model for PbTiO<sub>3</sub> so that it gives an out-of-plane polarization of  $1.0 \text{ C m}^{-2}$  at 0 K when subject to the epitaxial constraint just described. Care was needed because the model of ref. 24 for bulk PbTiO<sub>3</sub> becomes unstable when the epitaxial constraint is used in combination with the change in  $\epsilon_\infty$ . Nevertheless, it was possible to obtain a stable model with the correct ground-state polarization by adjusting the expansive hydrostatic pressure introduced in ref. 24 as an empirical correction: instead of the  $-13.9 \text{ GPa}$  used in ref. 24, here we used  $-11.2 \text{ GPa}$ . Also, when we use this model to simulate a film of PbTiO<sub>3</sub> under the SrTiO<sub>3</sub> epitaxial constraint, we get a ferroelectric transition temperature of 460 K, which is slightly below the temperature at which the fluctuating domains appear in the (8, 2) superlattice (490 K). As in the case of SrTiO<sub>3</sub>, the difference between bulk material and superlattice is probably caused by the different value of  $\epsilon_\infty$ : we use a slightly larger value for the pure film, which results in a weaker ferroelectric instability.

These approximations and adjustments allow us to construct models for superlattices of arbitrary ( $n_b, n_a$ ) stacking. For the simulations, we used periodically repeated supercells that contain  $10 \times 10$  elemental perovskite units in-plane, whereas out-of-plane they expand one full superlattice period. Thus, for example, for the (8, 2) superlattice, we used a simulation box that contains  $10 \times 10 \times (8 + 2) \times 5 = 5,000$  atoms. We solved the models by running Monte Carlo simulations comprising between 10,000 and 40,000 thermalization sweeps (longer thermalization is needed in the vicinity of phase transitions) followed by 50,000 sweeps to compute thermal averages. The dielectric susceptibility was calculated by applying a small out-of-plane electric field to the simulation box. We found that, in this highly reactive system, this approach converged much faster than the usual fluctuation formulas<sup>43</sup>.

The low-temperature ground state of our (8, 2) superlattice is sketched in Extended Data Fig. 2, in which the stripe domain structure can be nicely appreciated. This result closely resembles the one obtained directly from first-principles calculations<sup>44</sup> in the limit of 0 K; this agreement further confirms the accuracy of our model potential.

**Calculation of local dielectric constants.** In the following, we summarize the derivation of formulas that relate the local response of each layer with the global response of the superlattice. Here we are exclusively concerned with the response along the superlattice stacking direction. We use a '0' superscript to refer to the situation in which no external electric field is applied, and  $i$  to label the layers in the superlattice. In absence of free charges, the condition on the continuity of the electric displacement implies

$$D_i^0 = D^0 = P_i^0 + \epsilon_0 E_i^0$$

for all layers.  $E_i^0$  is the total electric field acting on layer  $i$ . In general, this total field can be split into local and external contributions, so that  $E_i = E_{i,\text{loc}} + E_{\text{ext}}$ . Naturally, when no external field is applied, we simply have  $E_i^0 = E_{i,\text{loc}}^0$ .

Additionally, if we have  $M$  layers in the repeated unit of the superlattice, the periodicity of the potential implies that

$$\sum_{i=1}^M l_i E_i^0 = 0$$

in which  $l_i$  is the thickness of layer  $i$ . Hence, in the absence of an applied field, there is no net potential drop across the supercell. Then, we immediately get that

$$D^0 = L^{-1} \sum_{i=1}^M l_i D_i^0 = L^{-1} \sum_{i=1}^M l_i P_i^0 = P^0$$

in which  $L = \sum_{i=1}^M l_i$  is one superlattice period and  $P^0$  is the polarization of the superlattice with no field applied. As a result, the electric field at layer  $i$  is

$$E_i^0 = E_{i,\text{loc}}^0 = (P^0 - P_i^0)/\epsilon_0$$

Now we consider an external electric field  $E_{\text{ext}}$ . It is trivial to verify that the field-induced variations in polarization, electric field and displacement satisfy

$$\Delta D_i = \Delta D = \Delta P_i + \epsilon_0(\Delta E_{i,\text{loc}} + E_{\text{ext}}) = \Delta P + \epsilon_0 E_{\text{ext}}$$

$$\sum_{i=1}^M l_i \Delta E_{i,\text{loc}} = 0$$

and

$$\Delta E_{i,\text{loc}} = (\Delta P - \Delta P_i)/\epsilon_0$$

Then, the dielectric constant of layer  $i$  is computed as

$$\epsilon_i = \frac{\Delta D}{\epsilon_0 \Delta E_i} = \frac{(\Delta P + \epsilon_0 E_{\text{ext}})}{\Delta P - \Delta P_i + \epsilon_0 E_{\text{ext}}} = \frac{(\chi + 1)}{\chi - \chi_i + 1} = \frac{\epsilon_{\text{tot}}}{\epsilon_{\text{tot}} - \chi_i}$$

where we have introduced the layer susceptibility

$$\chi_i = \frac{1}{\epsilon_0} \frac{\Delta P_i}{E_{\text{ext}}}$$

As a result, we have written all the relevant quantities in terms of the local susceptibilities  $\chi_i$ , which is convenient at conceptual and practical levels. Conceptually,  $\chi_i$  is a quantity we expect to be positive in all cases, because an applied electric field will create dipoles parallel to it. This basic local response of the material is physically and intuitively clear, because it is free from the subtleties (associated to the long-range electrostatic effects encapsulated in the local depolarizing fields) that affect the dielectric constant. Practically,  $\chi_i$  is very easy to compute from a Monte Carlo simulation, whether by explicitly applying an electric field and calculating the change in local polarization or by directly inspecting the fluctuations of the local polarizations in absence of applied field. The latter approach can be viewed as a generalization of the method described in, for example, ref. 43; similar fluctuation formulas for ferroelectric nanostructures were introduced in refs 13, 45.

The layers labelled by  $i$  will typically correspond to actual PbTiO<sub>3</sub> and SrTiO<sub>3</sub> layers, but we could also further sub-divide our superlattice. For example, the above formulas formally allow us to consider contributions from the interfaces, or from different regions within a layer. This kind of subdivision was used to prepare Fig. 3e, in which maps of the dielectric constant as a function of position along the superlattice stacking direction are reported.

The dielectric susceptibility  $\chi_i$  of a layer  $i$  can be viewed as a direct average of the susceptibilities  $\chi_i(x, y)$  coming from different regions of the  $x$ - $y$  plane of the layer. Hence, we can use a representation as in Fig. 3f to determine which part of a given layer (domain walls or domains) contributes the most to  $\chi_i$ . It could be tempting to interpret the layer dielectric constant  $\epsilon_i$  as coming from a collection of parallel capacitors, which would formally allow us to map  $\epsilon_i(x, y)$ . However, such a construction implicitly assumes an equal potential drop across the individual capacitors within layer  $i$ , which seems in conflict with the inhomogeneous in-plane structure of our PbTiO<sub>3</sub> layers.

For the calculation of local polarizations, we evaluated the local dipole and cell volume from the atomic positions and Born effective charges. We computed dipoles centred on the A (Pb/Sr) and B (Ti) sites of the perovskite structure, by considering the weighted contributions of the surrounding atoms. Thus, for example, the dipole centred on a specific Ti cation was computed by adding up contributions from the Ti itself, the six neighbouring oxygens (each such contribution was divided by two, because each oxygen has two first-neighbour Ti cations) and the eight neighbouring A (Pb/Sr) cations (each such contribution was divided by eight, as each A cation has eight first-neighbour Ti cations).

**Relation between atomistic simulations and experiment.** As already mentioned, our model potentials for PbTiO<sub>3</sub>-SrTiO<sub>3</sub> superlattices are not expected to render quantitatively accurate results. The difficulties in reproducing the behaviour of the bulk compounds in a quantitative way are discussed in refs 24, 46, in which evidence is given of the challenge these materials pose to first-principles methods. The model deficiencies are best captured by the error in the obtained transition temperatures: the model for PbTiO<sub>3</sub> used here gives a value of 440 K when solved in bulk-like conditions, far below the experimental result of 760 K. Similarly, we do not expect our models to accurately capture the dielectric response of the SrTiO<sub>3</sub> layers in the superlattice, which tend to be stiffer than the experimental ones. Fortunately, beyond these quantitative inaccuracies, the qualitative behaviour of individual PbTiO<sub>3</sub> and SrTiO<sub>3</sub> obtained from our simulations, as well as that of the PbTiO<sub>3</sub>-SrTiO<sub>3</sub> superlattice, seem perfectly in line with experimental observations and physical soundness.

With regard to the results for the PbTiO<sub>3</sub>-SrTiO<sub>3</sub> system, our atomistic simulations predict a phase transition occurring in two steps: at a relatively high temperature (490 K), the  $c/a$  ratio of the PbTiO<sub>3</sub> layers clearly reflects the onset of local instantaneous ferroelectric order; then, at a lower temperature (about 370 K), the static multidomain ferroelectric state freezes in. Thus, according to these simulations, the interval between 370 K and 490 K is characterized by strongly fluctuating ferroelectric domains. This result is likely to be affected by finite-size effects in our simulations; yet, given the very large separation of the two transition temperatures, and the easy and frequent domain rearrangements observed in our Monte Carlo simulations, we believe it should be taken seriously. Experimentally, preliminary measurements of the (5, 4)<sub>28</sub> PbTiO<sub>3</sub>-SrTiO<sub>3</sub> superlattice (Extended Data Fig. 6) indicate that the kink in the measured temperature dependence of the  $c/a$  ratio (usually assumed to mark the ferroelectric transition) occurs at a slightly higher

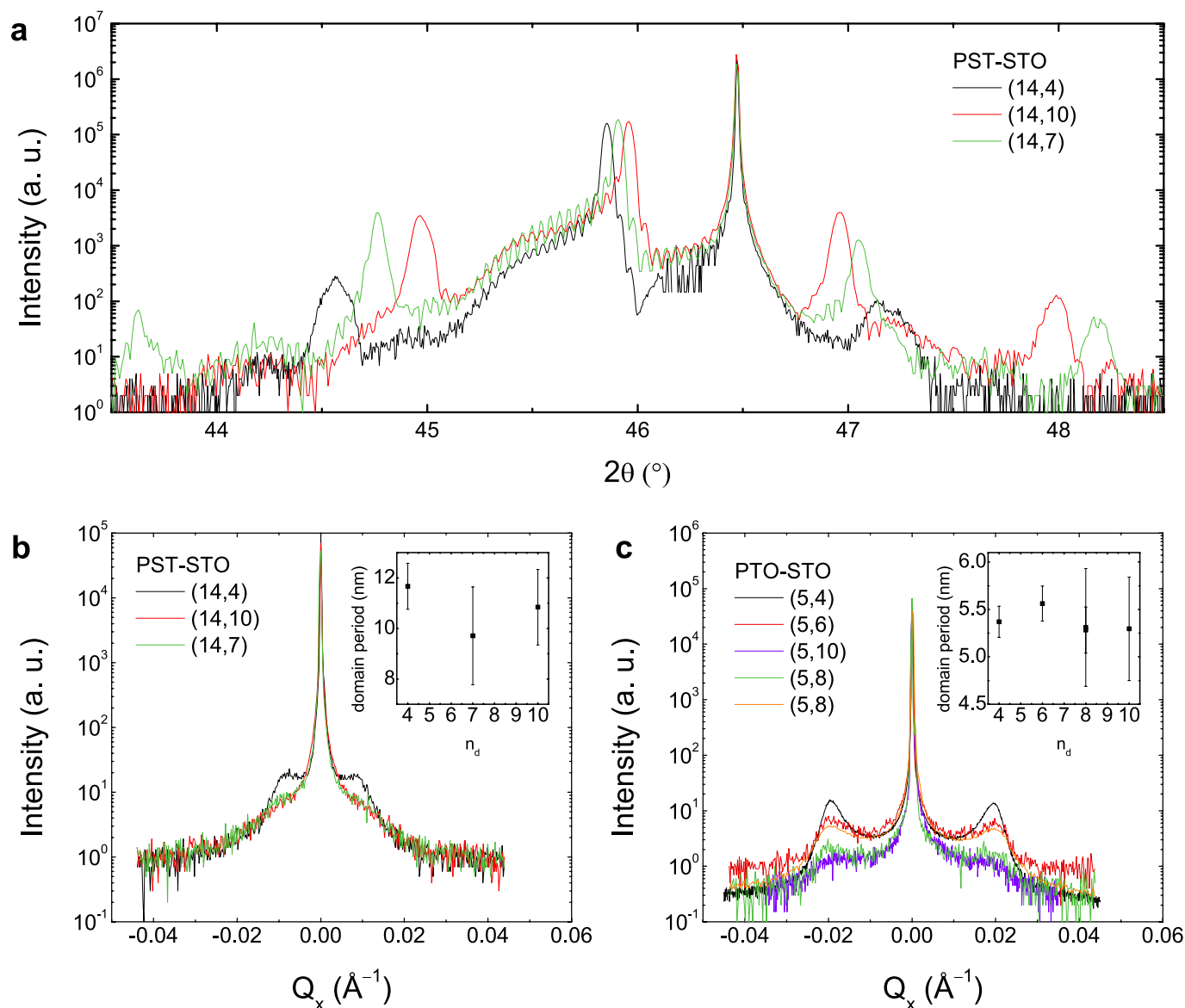
(about 50 K) temperature than the appearance of domain satellites in the diffuse scattering. A similar temperature difference has been noted in figure 3 of ref. 47 for a superlattice of a different composition. However, further investigation is needed to clarify the polarization structure in this temperature range.

We also ran simulations of various  $(8, n_d)$  superlattices, and computed the corresponding total dielectric constants as a function of temperature, to mimic our experimental approach to estimate the response of the  $\text{PbTiO}_3$  layer. Extended Data Fig. 3 shows the results for  $\epsilon_f$  obtained in this way: we find a temperature interval in which the  $\text{PbTiO}_3$  layers present a negative dielectric constant, which validates our experimental strategy for estimating  $\epsilon_f$ .

If we compare the results in Extended Data Fig. 3 and Fig. 3c, then we note that the temperature interval in which the negative capacitance is observed is essentially the same, but the quantitative values for  $\epsilon_f$  clearly differ. Nevertheless, given the approximations involved in each of the two methods to compute  $\epsilon_f$ —for example, heuristic division into  $\text{PbTiO}_3$  and  $\text{SrTiO}_3$  layers and the implicit assumption that  $\text{PbTiO}_3$  layers in  $(8, n_d)$  superlattices of varying  $\text{SrTiO}_3$  content behave equivalently—these quantitative discrepancies do not seem very substantial and we have not investigated them further.

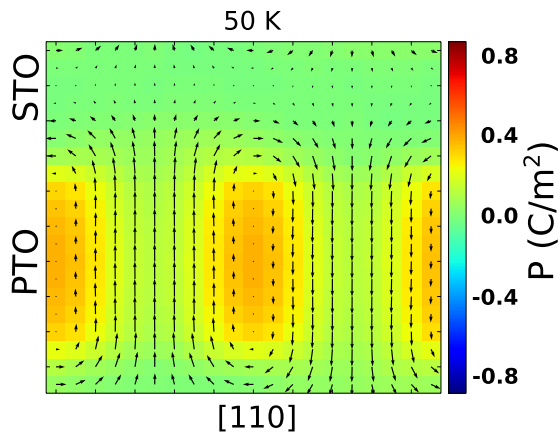
Finally, returning to our phenomenological predictions in Fig. 1c, it would appear that the Landau–Ginzburg result for the static domain structure bears a closer qualitative resemblance to the experimental data and the atomistic simulation results than does the Kittel model, despite the fact that it is the Kittel model that correctly captures the important contribution of domain-wall motion. However, in the Kittel model, we have not included the possibility of domain-wall pinning (by defects or otherwise), which would reduce the domain-wall contribution and could lead to an upturn of  $\epsilon_f$  at low temperatures. Quantifying the relative contributions to negative capacitance from domain-wall motion and the static domain response, both experimentally and through atomistic simulations, would be a worthwhile challenge for future studies.

31. Landau, L. & Lifshits, E. On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. *Phys. Zeitsch. Sow.* **8**, 153–169 (1935).
32. Kittel, C. Theory of the structure of ferromagnetic domains in films and small particles. *Phys. Rev.* **70**, 965–971 (1946).
33. Stephanovich, V. A., Luk'yanchuk, I. A. & Karkut, M. G. Domain-enhanced interlayer coupling in ferroelectric/paraelectric superlattices. *Phys. Rev. Lett.* **94**, 047601 (2005).
34. Catalan, G., Schilling, A., Scott, J. F. & Gregg, J. M. Domains in three-dimensional ferroelectric nanostructures: theory and experiment. *J. Phys. Condens. Matter* **19**, 132201 (2007).
35. Sené, A. *Theory of Domains and Nonuniform Textures in Ferroelectrics*. PhD thesis, Université de Picardie (2010).
36. Zubko, P. *et al.* Electrostatic coupling and local structural distortions at interfaces in ferroelectric/paraelectric superlattices. *Nano Lett.* **12**, 2846–2851 (2012).
37. Plonka, R., Dittmann, R., Pertsev, N. A., Vasco, E. & Waser, R. Impact of the top-electrode material on the permittivity of single-crystalline  $\text{Ba}_{0.7}\text{Sr}_{0.3}\text{TiO}_3$  thin films. *Appl. Phys. Lett.* **86**, 202908 (2005).
38. Stengel, M. & Spaldin, N. A. Origin of the dielectric dead layer in nanoscale capacitors. *Nature* **443**, 679–682 (2006).
39. Catalan, G., O'Neill, D., Bowman, R. M. & Gregg, J. M. Relaxor features in ferroelectric superlattices: a Maxwell–Wagner approach. *Appl. Phys. Lett.* **77**, 3078–3080 (2000).
40. Ghosez, P., Cockayne, E., Waghmare, U. V. & Rabe, K. M. Lattice dynamics of  $\text{BaTiO}_3$ ,  $\text{PbTiO}_3$ , and  $\text{PbZrO}_3$ : a comparative first-principles study. *Phys. Rev. B* **60**, 836–843 (1999).
41. Bousquet, E. *et al.* Improper ferroelectricity in perovskite oxide artificial superlattices. *Nature* **452**, 732–736 (2008).
42. Lisenkov, S. & Bellaiche, L. Phase diagrams of  $\text{BaTiO}_3/\text{SrTiO}_3$  superlattices from first principles. *Phys. Rev. B* **76**, 020102 (2007).
43. García, A. & Vanderbilt, D. Electromechanical behavior of  $\text{BaTiO}_3$  from first principles. *Appl. Phys. Lett.* **72**, 2981–2983 (1998).
44. Aguado-Puente, P. & Junquera, J. Structural and energetic properties of domains in  $\text{PbTiO}_3/\text{SrTiO}_3$  superlattices from first principles. *Phys. Rev. B* **85**, 184105 (2012).
45. Ponomareva, I., Bellaiche, L. & Resta, R. Relation between dielectric responses and polarization fluctuations in ferroelectric nanostructures. *Phys. Rev. B* **76**, 235403 (2007).
46. Wojdel, J. C. & Íñiguez, J. Testing simple predictors for the temperature of a structural phase transition. *Phys. Rev. B* **90**, 014105 (2014).
47. Zubko, P. *et al.* Ferroelectric domains in  $\text{PbTiO}_3/\text{SrTiO}_3$  superlattices. *Ferroelectrics* **433**, 127–137 (2012).

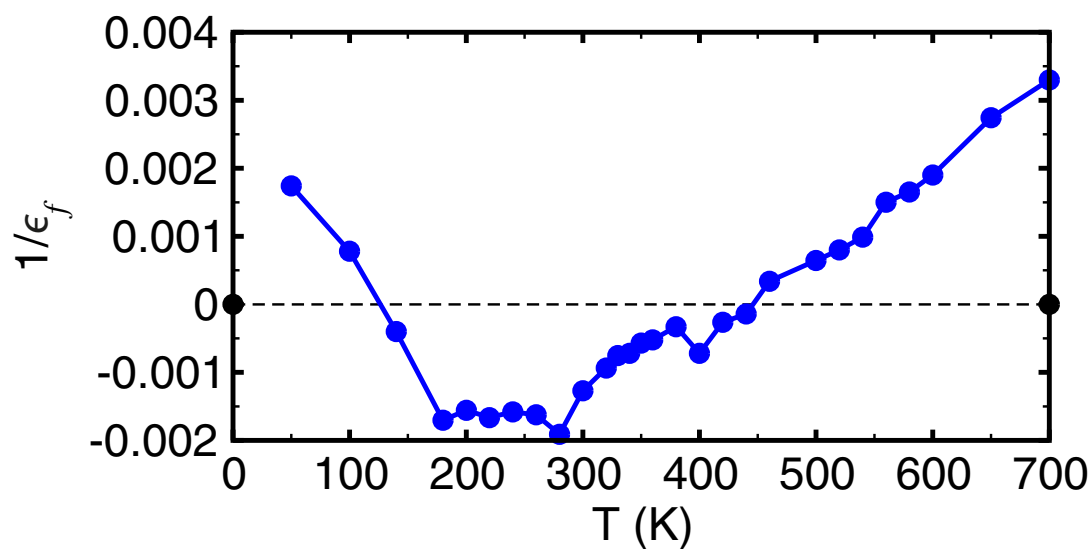


**Extended Data Figure 1 | XRD characterization of the superlattices.** **a**, Intensity profiles around the (002) substrate reflection for  $\text{Pb}_{0.5}\text{Sr}_{0.5}\text{TiO}_3$ - $\text{SrTiO}_3$  (PST-STO) superlattices. The broad peaks around  $2\theta = 45.5^\circ$  correspond to the top and bottom  $\text{SrRuO}_3$  electrodes. Finite-size oscillations due to the 200-nm superlattice thickness are visible. **b**, **c**, XRD domain satellites for  $\text{Pb}_{0.5}\text{Sr}_{0.5}\text{TiO}_3$ - $\text{SrTiO}_3$  (**b**) and  $\text{PbTiO}_3$ - $\text{SrTiO}_3$  (PTO-STO; **c**) superlattices. Insets, domain periodicities

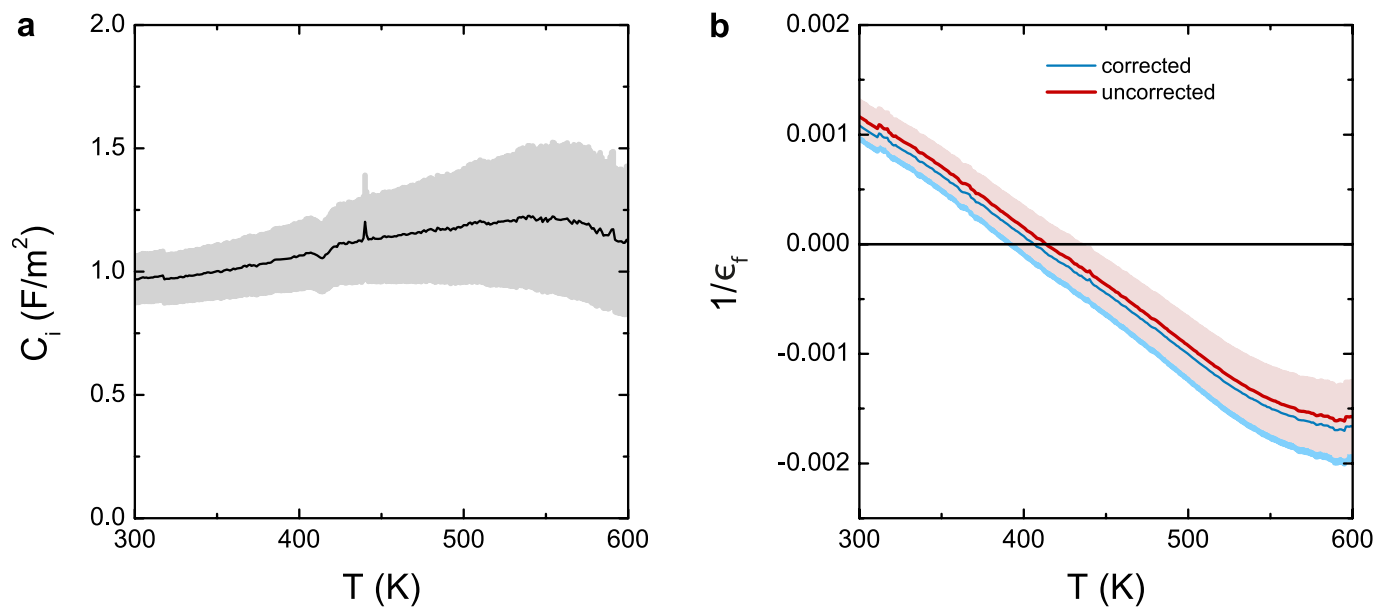
obtained from fitting the  $Q_x$  line profiles using a sum of two Gaussian functions for the domain satellites and a Lorentzian function for central Bragg peak. The error bars were determined from the 95% confidence bounds for the peak positions obtained from the fits.  $n_d$  is the number of  $\text{SrTiO}_3$  layers in the (14,  $n_d$ ) (**b**) or (5,  $n_d$ ) (**c**) superlattices;  $Q_x$  is the in-plane reciprocal-space coordinate.



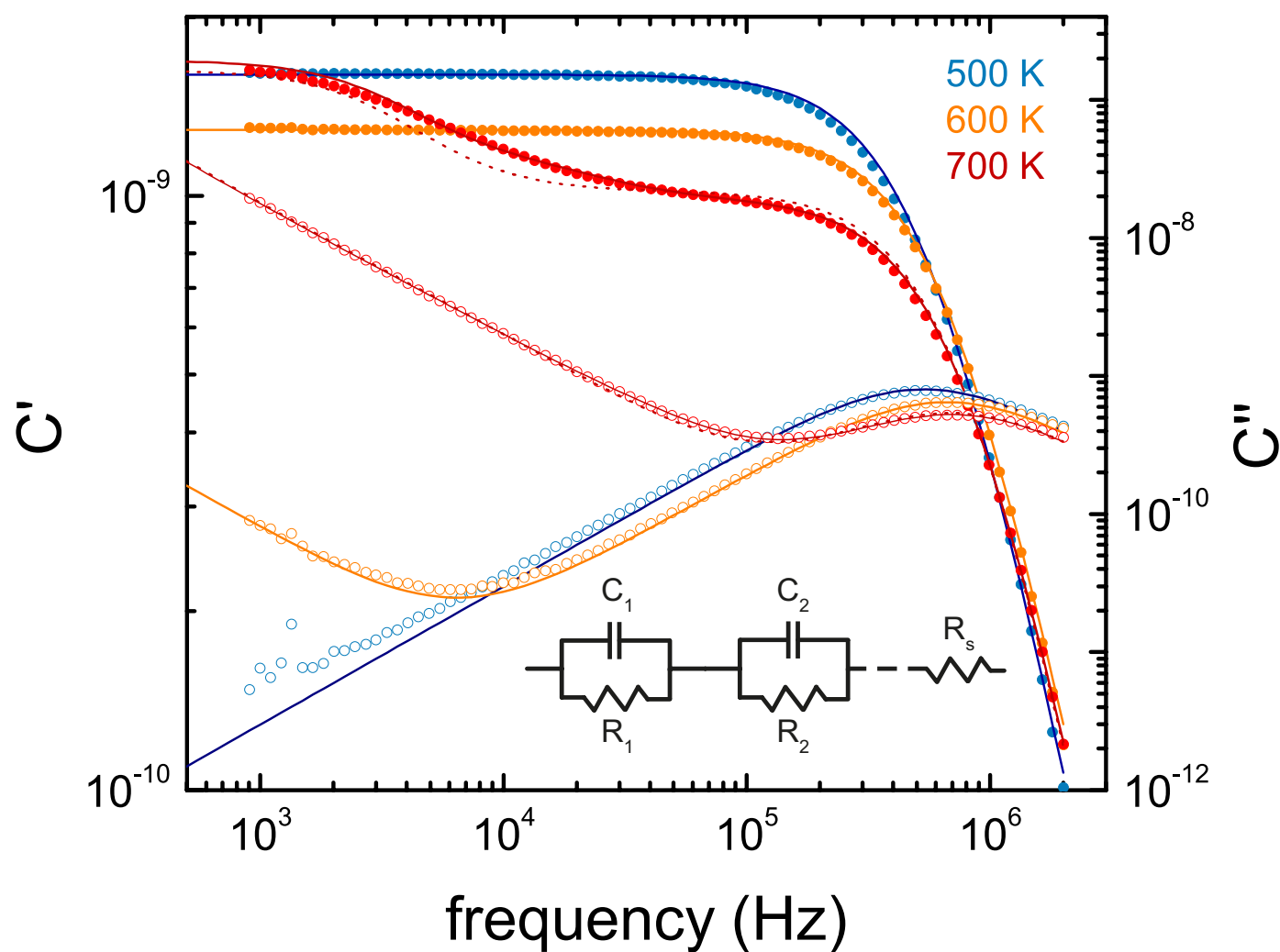
**Extended Data Figure 2 | Local polarization distribution at low temperature.** Arrows indicate the dipole component within the  $(\bar{1}10)$  plane; we plot arrows for Pb/Sr-centred and Ti-centred dipoles. The colouring indicates the polarization  $P$  component along  $[\bar{1}10]$ , revealing a low-temperature polar order at the domain walls. PTO,  $\text{PbTiO}_3$ ; STO,  $\text{SrTiO}_3$ .



**Extended Data Figure 3 | Comparison with experiment.** Reciprocal dielectric constant  $1/\epsilon_f$  of the  $\text{PbTiO}_3$  layers as a function of temperature  $T$ , calculated from the computed total dielectric constants of  $(8, n_d)$  superlattices using the same analysis as for the experimental data.

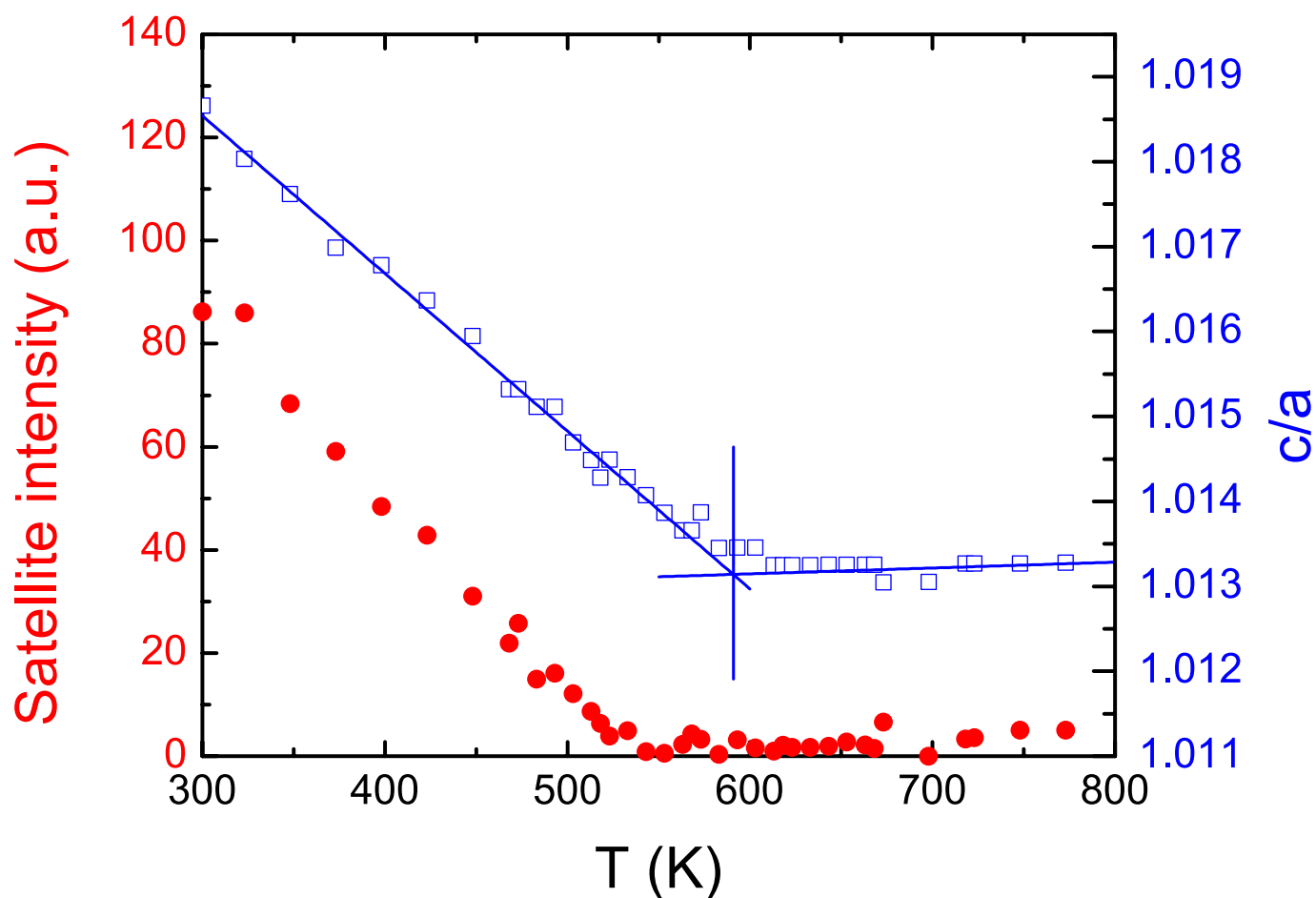


**Extended Data Figure 4 | Interface capacitance contributions.** **a**,  $\text{SrRuO}_3$ – $\text{SrTiO}_3$  interface contribution  $C_i$  to the dielectric response. **b**, Dielectric stiffness  $1/\epsilon_f$  of the  $\text{PbTiO}_3$  layers with (blue) and without (red) correcting for the interface capacitance. Grey (**a**) and blue and pink (**b**) shading indicates estimated uncertainties obtained from weighted-least-squares linear fits.



**Extended Data Figure 5 | Dielectric impedance spectroscopy of  $\text{PbTiO}_3\text{-SrTiO}_3$  superlattices.** Real ( $C'$ ; filled circles) and imaginary ( $C''$ ; open circles) parts of the complex capacitance function  $C = C' + iC''$  for a  $(5, 8)_{30}$   $\text{PbTiO}_3\text{-SrTiO}_3$  superlattice. For temperatures below about 650 K, the data are well fitted by a single parallel R-C element in series with  $R_s$ , as shown by solid curves for the 500 K (blue) and 600 K (orange) data. At higher temperatures, Maxwell-Wagner relaxations appear as the

conductivities of some layers increase faster with temperature than others. At 700 K (red), the response is qualitatively captured by a model with two parallel R-C elements in series with each other (dashed red curve), whereas for a quantitative fit three R-C elements are required (solid red curve). The inset shows the arrangement of elements in the generalized equivalent circuit used to fit the data.



**Extended Data Figure 6 | Temperature evolution of the tetragonality and domain satellites.** Intensity of the XRD domain satellite (filled red circles) and the film tetragonality ( $c/a$ ; open blue squares) for a (5, 4)<sub>28</sub> PbTiO<sub>3</sub>-SrTiO<sub>3</sub> superlattice. The satellite intensity was obtained by

integrating the measured intensity of the domain satellites and subtracting the minimum integrated intensity in the paraelectric phase. Vertical blue line marks the temperature at which linear fits to the low- and high-temperature data (blue lines) intersect.

# Controlled fragmentation of multimaterial fibres and films via polymer cold-drawing

Soroush Shabahang<sup>1</sup>, Guangming Tao<sup>1</sup>, Joshua J. Kaufman<sup>1</sup>, Yangyang Qiao<sup>2</sup>, Lei Wei<sup>3</sup>, Thomas Bouchenot<sup>2</sup>, Ali P. Gordon<sup>2</sup>, Yoel Fink<sup>4</sup>, Yuanli Bai<sup>2</sup>, Robert S. Hoy<sup>5</sup> & Ayman F. Abouraddy<sup>1,6</sup>

Polymer cold-drawing<sup>1–4</sup> is a process in which tensile stress reduces the diameter of a drawn fibre (or thickness of a drawn film) and orients the polymeric chains. Cold-drawing has long been used in industrial applications<sup>5–7</sup>, including the production of flexible fibres with high tensile strength such as polyester and nylon<sup>8,9</sup>. However, cold-drawing of a composite structure has been less studied. Here we show that in a multimaterial fibre<sup>10,11</sup> composed of a brittle core embedded in a ductile polymer cladding, cold-drawing results in a surprising phenomenon: controllable and sequential fragmentation of the core to produce uniformly sized rods along metres of fibre, rather than the expected random or chaotic fragmentation. These embedded structures arise from mechanical–geometric instabilities associated with ‘neck’ propagation<sup>2,3</sup>. Embedded, structured multimaterial threads with complex transverse geometry are thus fragmented into a periodic train of rods held stationary in the polymer cladding. These rods can then be easily extracted via selective dissolution of the cladding, or can self-heal by thermal restoration to re-form the brittle thread. Our method is also applicable to composites with flat rather than cylindrical geometries, in which case cold-drawing leads to the break-up of an embedded or coated brittle film into narrow parallel strips that are aligned normally to the drawing axis. A range of materials was explored to establish the universality of this effect, including silicon, germanium, gold, glasses, silk, polystyrene, biodegradable polymers and ice. We observe, and verify through nonlinear finite-element simulations, a linear relationship between the smallest transverse scale and the longitudinal break-up period. These results may lead to the development of dynamical and thermoreversible camouflaging via a nanoscale Venetian-blind effect, and the fabrication of large-area structured surfaces that facilitate high-sensitivity bio-detection.

When a longitudinal tensile stress is applied to a ductile polymer fibre or sheet, the mechanical instability known as ‘necking’ reduces the transverse dimensions of the sample and longitudinally orients the constituent chains<sup>2–4</sup>. The necked region is initially localized, but expands uniformly via propagation of the ‘shoulder’ into undeformed regions of the polymer until the neck is fully developed and extends throughout the sample length (Fig. 1). The polymer is consequently left in a new, anisotropic phase with potentially superior mechanical and/or optical properties<sup>1,5</sup>. For example, in the original observation of cold-drawing<sup>1</sup>, a highly oriented, transparent and robust neck developed in an initially opaque and brittle polyester fibre. Commercial applications that exploited the large tensile strength, low weight and high flexibility of cold-drawn synthetic polymer fibres soon followed<sup>5–7</sup>. After decades of experimental and theoretical research<sup>1–9</sup>, cold-drawing is now reasonably well-understood at a phenomenological level.

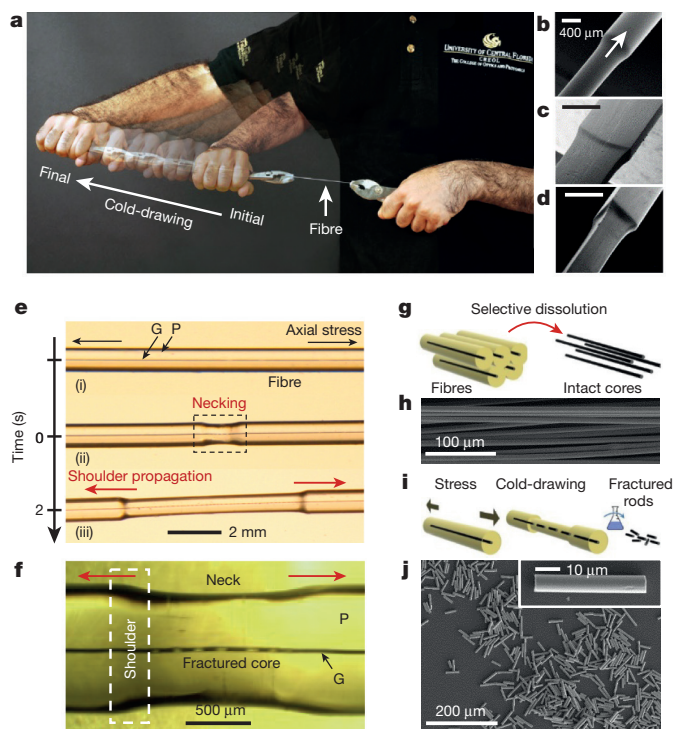
To date, investigations of cold-drawing have focused on the changes that occur in monolithic structures (composed of a single material)

upon propagation of the shoulder, as well as further deformation leading to material failure<sup>4</sup>. Studies of the cold-drawing of polymer composites have primarily concentrated on improving the mechanical properties of bulk materials via fibre reinforcement<sup>12–14</sup>. Conversely, the study of fibre-reinforced composites has concentrated on fracture phenomena driven by stress-transfer in a matrix<sup>15–17</sup>. Here we report a new dynamical phenomenon that exploits cold-drawing in multimaterial composite structures—in the form-factor of a cylindrical fibre or a flat film (Fig. 1b–d)—and combines disparate materials, only one of which (a thermoplastic polymer) is amenable to cold-drawing. The mechanical–geometric mismatch between the ductile polymer (which undergoes cold-drawing) and the other, relatively brittle material (which does not) is harnessed as a scalable, mechanical pathway to produce a wide variety of complex multi-component nano- and microstructures with arbitrary cross-sectional geometries through the controllable local fragmentation of the brittle materials within the axially propagating polymer shoulders.

We start by describing our observations in the context of the cylindrical multimaterial fibre shown in Fig. 1e, f. The fibre consists of a 20- $\mu\text{m}$ -diameter glass core (the inorganic chalcogenide glass  $\text{As}_2\text{Se}_3$ ) embedded in a 1-mm-diameter polymer cladding (the thermoplastic polymer polyethersulfone, PES)<sup>10,11</sup>. Fibres can be tens of metres in length; see Supplementary Information for fibre fabrication. The use of PES is not required; we have reproduced our results in fibres made of other thermoplastic polymers, including polycarbonate, polyetherimide, polysulfone and cyclic olefin polymer. At room temperature, the core is brittle whereas the polymer is ductile. At a homogeneous uniaxial extension of a few per cent, necks form locally and extend until the fibre is fully drawn. Videos capturing the neck expansion and shoulder propagation along the fibre in real time reveal a surprising dynamical phenomenon that takes place inside the polymer shoulders during their propagation along the fibre (Supplementary Video 1). Although the core is initially intact along the fibre axis, within the propagating shoulder the glassy core fragments in an orderly sequence (upon passage of the shoulder) into a periodic train of cylindrical rods that are held stationary in the fibre and separated by voids. As the shoulders propagate further, they continue to fragment the core *in situ* until they consume the whole length of the fibre, or until the applied stress is removed. We harvest the glass micro-rods by selectively dissolving the polymer with an organic solvent (dimethylacetamide, DMAC; Fig. 1g–j); the rough faceted surfaces confirm that the rods result from brittle fracture (Fig. 1j, inset).

This phenomenon is reminiscent of shear-lag fracture (SLF) in fibre-reinforced composites with low interfacial strength<sup>15–17</sup>. SLF is a quite general phenomenon occurring in composites of mechanically incompatible materials. Previous work has generally sought to suppress SLF as a means of increasing the ductility of fibres, whereas here we have induced a controlled form of SLF at the neck front.

<sup>1</sup>CREOL, The College of Optics & Photonics, University of Central Florida, Orlando, Florida 32816, USA. <sup>2</sup>Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, Florida 32816, USA. <sup>3</sup>School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore. <sup>4</sup>Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Department of Physics, University of South Florida, Tampa, Florida 33620, USA. <sup>6</sup>Department of Materials Science and Engineering, University of Central Florida, Orlando, Florida 32816, USA.



**Figure 1 | Fragmentation via a cold-drawing-induced, propagating, mechanical-geometric instability.** **a**, Photograph of a polymer fibre undergoing cold-drawing under axial stress at a speed of approximately  $5 \text{ mm s}^{-1}$ . Multiple shots taken over 1 min are overlaid to highlight the extent of fibre elongation. **b–d**, Scanning electron microscope (SEM) micrographs of the propagating shoulder in polymer (PES) fibres with cross-sections that are circular (diameter of  $0.7 \text{ mm}$ ; **b**), rectangular (side lengths of  $0.2 \text{ mm}$  and  $1 \text{ mm}$ ; **c**) and equilaterally triangular (side length of  $0.4 \text{ mm}$ ; **d**). Scale bars,  $400 \mu\text{m}$ . **e**, Transmission optical micrographs of a multimaterial cylindrical fibre undergoing cold-drawing at  $3 \text{ mm s}^{-1}$  captured at three different stages: (i) initially intact fibre; (ii) neck formation; and (iii) shoulder propagation, leaving behind a fractured core after fragmentation. The cladding is a polymer 'P' (PES) and the core is a glass 'G' ( $\text{As}_2\text{Se}_3$ ). **f**, A magnified transmission micrograph of the neck region, corresponding to the dashed black rectangle in **e**. The dashed white rectangle highlights the propagating instability, wherein fragmentation takes place. **g**, Schematic of selective dissolution of the polymer cladding to retrieve intact cores. **h**, SEM micrograph of retrieved intact glass cores from multiple fibres. **i, j**, Schematic (**i**) and SEM micrograph (**j**) of retrieved nano-fragmented micro-rods by selective dissolution from a cold-drawn fibre. Inset in **j** is an SEM micrograph of a single micro-rod.

Although a 'local' brittle fracture takes place, as in traditional SLF studies, the global dynamics are different; see Extended Data Fig. 1. We have found that this phenomenon applies to many core materials. These materials include dense solids, such as crystalline semiconductors (Si and Ge)<sup>18</sup>, and inorganic glasses (such as silicates, phosphates, chalcogenides and tellurites)<sup>19</sup>. Relatively brittle polymers that do not typically undergo cold-drawing themselves can also be fragmented in this manner—including polystyrene, the biodegradable polymer polyethylene oxide, and even natural polymers such as silk and human hair. Indeed, even ice inside a hollow-core polymer fibre undergoes a similar break-up process before melting (Supplementary Fig. S6).

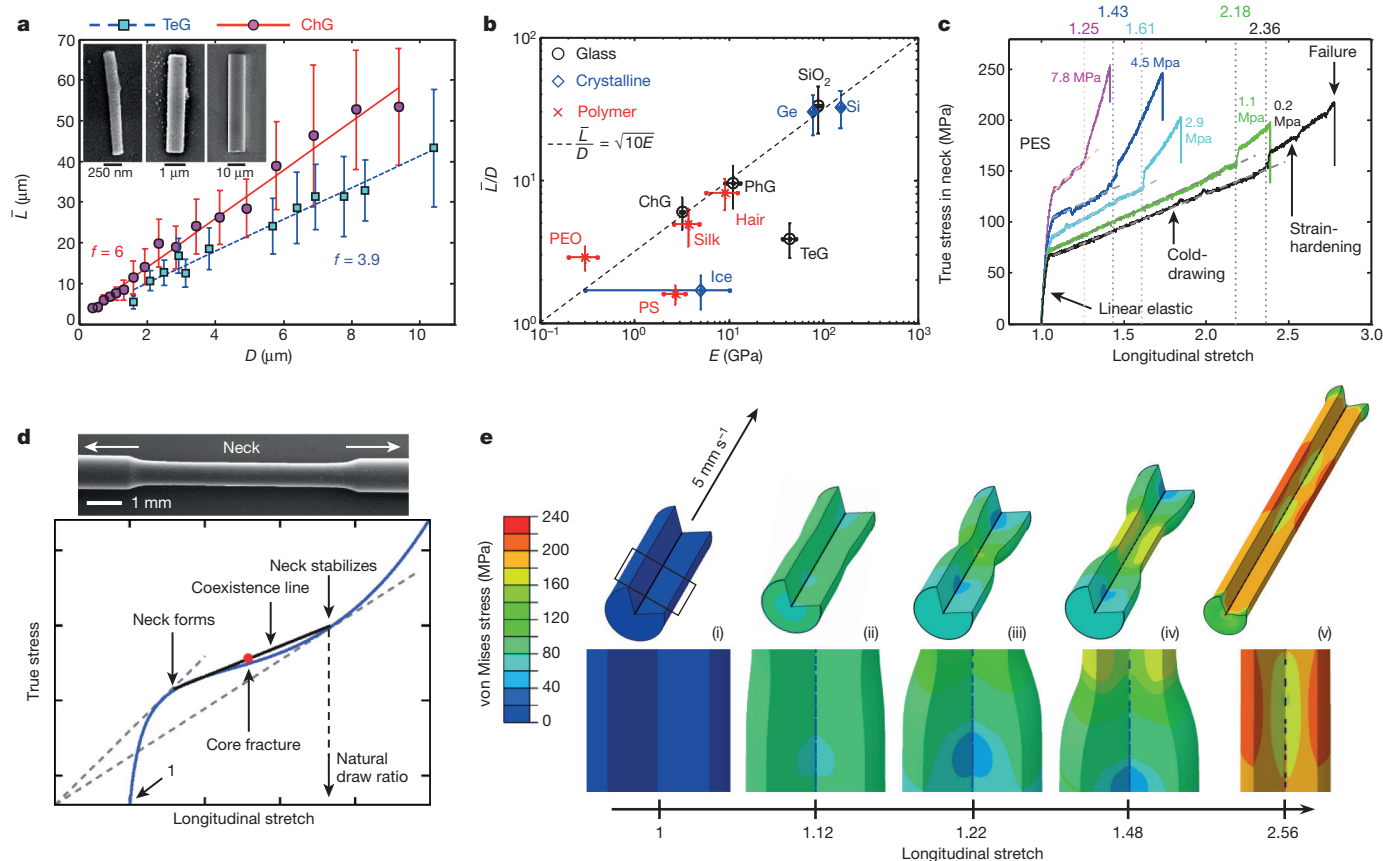
Systematic investigation leads to several non-intuitive findings. First, the average length  $\bar{L}$  of the glass rods is linearly proportional to the core diameter  $D$ , as shown in Fig. 2a. Second, the linear proportionality  $\bar{L}/D = f$  depends on the core material, but not on the details of the experimental procedure (such as drawing speed, pre-stress of the polymer cladding or fibre outer diameter). Specifically, although the values of the strain at which necking is initiated and the final drawing ratio both depend on the fabrication conditions of the fibre, such

as the tension under which it is thermally drawn from the melt (Fig. 2c), they do not affect  $f$ . Note that the linear  $\bar{L}/D$  proportionality is 'cut off' at an upper limit set by the length of the shoulder region  $L_{\text{shoulder}}$ : it requires that  $\bar{L}, D \ll L_{\text{shoulder}}, D_{\text{cladding}}$ , where  $D_{\text{cladding}}$  is the diameter of the fibre cladding. Third, by determining the ratio  $f$  for a wide range of materials (Fig. 2b and Extended Data Table 1), we find empirically that  $f \propto \sqrt{E}$ , where  $E$  is the Young's modulus of the core; for example,  $\text{As}_2\text{Se}_3$  cores always have  $f \approx 6$ , whereas tellurite glass cores have  $f \approx 3.9$  (Fig. 2a). Analytic treatments of SLF vary in sophistication from Cox's original heuristic treatment<sup>20</sup> to Nairn's rigorous approach<sup>15</sup>, but agree that  $\bar{L}/D = \sqrt{E/\Omega}$ , consistent with Fig. 2b. Here  $\Omega$  is a characteristic stress or energy density; for example,  $\Omega$  is the Young's modulus of the cladding in Cox's theory, whereas later theories yield more complicated expressions for  $\Omega$ . The data we show in Fig. 2b indicate that  $\Omega \approx 0.1 \text{ GPa}$  in our experiments, and the data in Fig. 2c reveal that the true stress within the neck is also approximately  $0.1 \text{ GPa}$  (our key stress scale).

These results are non-intuitive because stress distributions in the region in which the shoulder meets un-drawn material are known to be complex<sup>21</sup>, and the presence of the core introduces additional strain localization due to the mechanical incompatibility between the core and cladding. Some essential features may be captured using a simple Considère model for core fracture in the shoulder region (Fig. 2d). The fragmentation of glass cores may be readily understood as arising from a mechanical-geometric instability. The Considère criterion for necking of solid materials states that undeformed and necked regions coexist. Within the shoulder, complex gradients of these quantities are present, but the brittle fracture of glass cores occurs along the 'coexistence' line shown in Fig. 2d. The specific location along this line at which fracture occurs is material-dependent and sets the ratio  $f$ .

Given the above-mentioned complexity of the stress distributions, and to better understand the experimental findings, we perform non-linear finite-element simulations of the cold-drawing process for a typical core/cladding ( $\text{As}_2\text{Se}_3/\text{PES}$ ) cylindrical geometry; see Methods. Axisymmetric elements are used and the measured mechanical properties (including elastic modulus, plastic hardening and fracture) of the core and cladding are input to the computational model (Extended Data Fig. 2). The same tensile boundary conditions used in the experiments are applied. The simulation results are illustrated in Fig. 2e with the contour plots of von Mises stress distributions as a function of stretch. In step (ii), necking and core-fracture initiate, with the latter associated with the onset of stress concentration. Necking continues and the shoulders propagate in steps (iii) and (iv), resulting in sequential fragmentation of the core with a length-to-diameter ratio  $f$  of the rods that agrees with the experiments (Extended Data Table 2). As the rods pass into the necked region, the increased stretch expands the voids separating the fragments. The final fully drawn configuration is shown in step (v); see Supplementary Video 2 for a video of the simulation, and Extended Data Fig. 1 for a comparison with the more typical stress-transfer fragmentation in fibre-reinforced composites.

As we have demonstrated through cold-drawing of fibres with rectangular and triangular cross-sections (Fig. 1c, d), the phenomena associated with cold-drawing in multimaterial fibres are not restricted to cylindrical geometries. In the flat-fibre geometry, new in-fibre fragmentation phenomena are observed. In particular, when a thin brittle film is embedded in such a fibre (Fig. 3a, b), propagation of the rectangular shoulder upon necking leads progressively to fragmentation of the film into well-ordered strips (Supplementary Video 3). Comparable fracture was reported in ref. 22, in which deformation of metal films on polymer substrates was examined and strategies for maximizing the ductility of these composites were explored. As seen in Fig. 3c, the straight, sharp-edged strips resulting from a glass film fragmenting extend across the width of the fibre (about  $1 \text{ mm}$ ) and are separated by rectangular voids in a well-ordered array extending along the whole fibre length. The optical properties of the flat fibre change markedly as a result of cold-drawing. The undrawn fibre contains



**Figure 2 | Characterization of fragmentation induced by cold-drawing of a PES fibre.** **a**, Measurements of the average length  $\bar{L}$  of fragmented micro- and nano-rods of chalcogenide glass (ChG; As<sub>2</sub>Se<sub>3</sub>; red circles) and tellurite glass (TeG; 70TeO<sub>2</sub>–20ZnO–5K<sub>2</sub>O–5Na<sub>2</sub>O; blue squares) of diameter  $D$  in a PES fibre upon cold-drawing. The red solid and blue dashed lines are linear fits with slopes  $f = \bar{L}/D = 6$  and  $3.9$  for ChG and TeG, respectively. Vertical error bars represent the root-mean-squared (r.m.s.) length dispersion of rods at each value of  $D$  (Supplementary Information). Insets are SEM micrographs of individual rods resulting from the cold-drawing-driven fragmentation of As<sub>2</sub>Se<sub>3</sub> cores of diameters (from left to right) 200 nm, 1 μm and 10 μm. **b**, Measured values of  $f$  for a host of materials embedded in a PES fibre plotted against their Young's modulus  $E$ ; see Extended Data Table 1. The dashed line corresponds to the ansatz  $f \approx \sqrt{E/\Omega}$ , with  $\Omega = 0.1$  GPa (such that  $f \approx \sqrt{10E}$  when  $E$  is in gigapascals). Vertical error bars represent the measured r.m.s. dispersion in  $f$ ; horizontal error bars correspond to the uncertainty in the measured  $E$  (those for TeG and ice<sup>29</sup> reflect the range of reported values). PhG, phosphate glass; PEO, polyethylene oxide; PS, polystyrene. **c**, Stress–strain measurements of cylindrical PES fibres

a 300-nm-thick, continuous As<sub>2</sub>Se<sub>3</sub> film (Fig. 3b), which renders the fibre opaque. Fragmentation (Fig. 3c) reduces the optical opacity as a result of the voids opening up between the strips in the transparent polymer (Fig. 3d, e). Optical spectral transmission measurements reveal a blueshift of approximately 300 nm in the wavelength of the absorption edge after film fragmentation (Fig. 3d). These observations are compared to theoretical predictions for a Fabry–Pérot optical model of the fibre that additionally takes into account absorption, diffraction and Fresnel reflection<sup>23</sup> (Fig. 3e); see Methods. It is important to note that the change in the optical properties upon cold-drawing is a consequence of the mechanical–geometric transformation undergone by the embedded structure and not of the polymer itself.

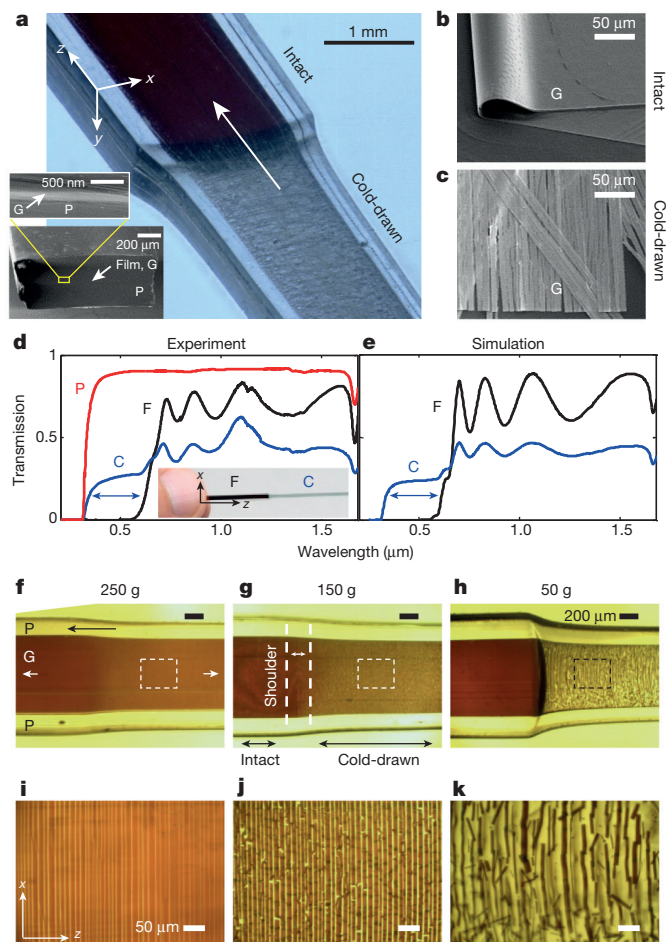
Figure 3f–k reveals the effect of polymer pre-stress (applied during thermal drawing of the fibre from the melt) on the uniformity and integrity of the fragmented strips. Large pre-stresses (Fig. 3f, i) produce smaller draw ratios<sup>24</sup> (less transverse contraction during neck

produced by thermal drawing<sup>30</sup> at different pre-stress values (ranging from 0.2 MPa to 7.8 MPa; see coloured-coded labels) identifying the four stages of linear elasticity, cold-drawing, strain-hardening and failure. The coexistence (dashed) lines and natural draw ratios (vertical dotted lines; values given above the plot) at neck stabilization (both defined in d) are identified.

**d**, Schematic representation of the Considère model. The blue curve corresponds to the true stress versus stretch in a strain-controlled experiment. The solid black ‘coexistence’ line indicates necked and unnecked regions coexisting at equal engineering stresses as local stretch varies from the onset of necking to the natural draw ratio at which stable neck propagation occurs. The dashed black lines serve as guides to the eye. Above the plot is an SEM micrograph of a necked region in a PES fibre; the arrows indicate the direction of the axial stress and shoulder propagation. **e**, von Mises stress distributions from finite-element simulations of in-fibre core (As<sub>2</sub>Se<sub>3</sub>) fragmentation during cold-drawing (at 5 mm s<sup>-1</sup>) of a PES fibre. The five steps (i)–(v) correspond to increasing stretch values. Top panels depict the full fibre; bottom panels show the regions corresponding to that highlighted by the rectangle in (i).

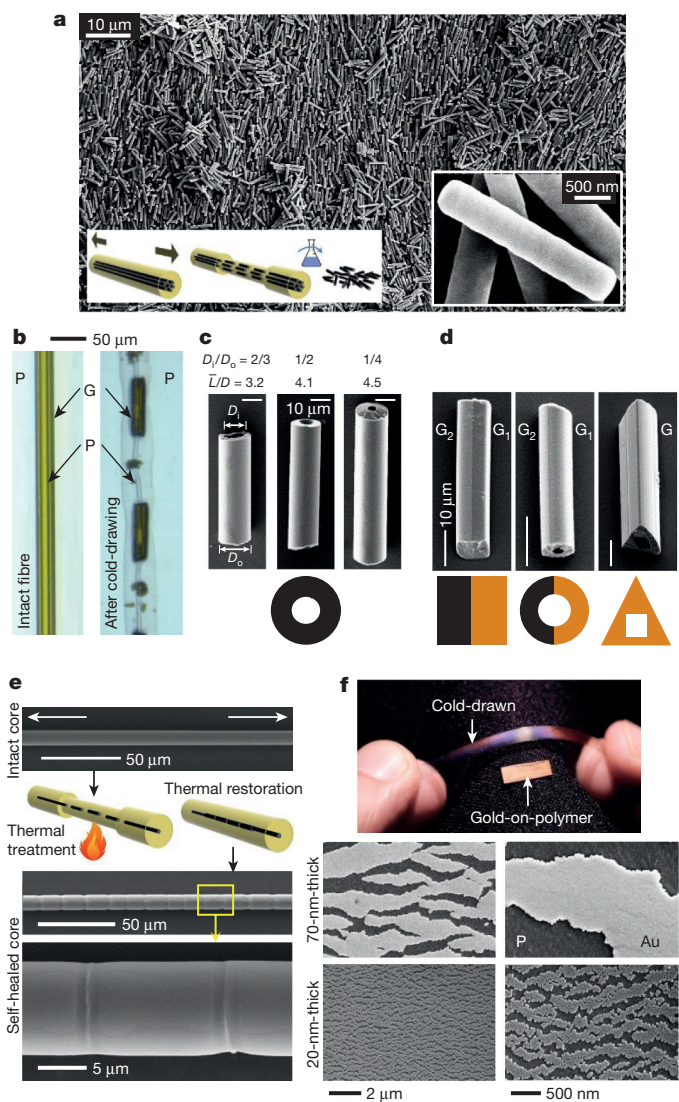
propagation) and a more uniform stress field within the shoulder, resulting in longer and more-parallel strips. This effect is a consequence of the greater orientation of polymer chains along the thermal drawing direction when compared to fibres drawn at lower pre-stress values. Low pre-stresses result in opposite trends; the film fragmentation is more violent because the strain field is more two-dimensional; that is, both the axial and transverse components are non-negligible. Shear bands in low-melt pre-stressed fibres are expected to play a substantial part in this outcome, which is borne out in finite-element simulations; see Methods, Extended Data Fig. 3, and Supplementary Video 4 for a video of the simulation. We retrieve in our simulations the experimentally observed ratio of strip width to film thickness (Fig. 3f–k) and also observe prominent shear-banding in the polymer that could account for the more violent fragmentation in the case of low pre-stress.

Several unique features of the in-fibre fragmentation process are highlighted in Fig. 4. First, the procedure extends to scenarios in which



**Figure 3 | Fragmentation of a thin film embedded in a flat polymer fibre undergoing cold-drawing.** **a**, Optical micrograph of a flat PES ('P') fibre containing a 300-nm-thick film of  $\text{As}_2\text{Se}_3$  ('G') (cross-section shown in inset), cold-drawn under axial stress. The propagating flat shoulder in the direction of the arrow separates an intact region (darker colour) and a reduced-size cold-drawn region (lighter colour). **b**, SEM micrograph of intact glass ('G') film retrieved from the fibre before cold-drawing. **c**, SEM micrograph of fragmented glass strips retrieved from the fibre after cold-drawing. **d**, **e**, Measured (**d**) and simulated (**e**) optical transmission through the flat fibre before ('F') and after ('C') cold-drawing. For reference, the transmission through the polymer alone ('P') is plotted in **d**. Inset is a photograph of a section of a fibre showing the change in the optical appearance after cold-drawing-driven fragmentation. **f-h**, Optical transmission micrographs of the necking region for three fibres produced at different pre-stress levels (250 g, **f**; 150 g, **g**; 50 g, **h**). The white arrows indicate the axial stress applied and the back arrow indicates the shoulder-propagation direction. Scale bars, 200 µm. **i-k**, Optical transmission micrographs of the sections in **f-h** enclosed in dashed boxes. Scale bars, 50 µm.

a large number of cores are embedded in a single polymer fibre<sup>25</sup> (about 4,000 cores, each of 500-nm diameter; Fig. 4a). All of the cores simultaneously undergo fragmentation into uniformly sized rods. Thus, this process is readily scalable to large-volume production. Supplementary Video 5 shows an example of a parallelepiped polymer fibre containing six cylindrical  $\text{As}_2\text{Se}_3$  cores all undergoing fragmentation upon passage of the shoulder. Second, the length-to-diameter ratio may be tuned by using cylindrical shells rather than solid cores (Fig. 4b). Indeed,  $\bar{L}/D$  for  $\text{As}_2\text{Se}_3$ , for example, changes from 6 for a solid core (Fig. 2a) to 4.5 and 3.2 when the ratio of the outer to inner diameter of the shell is 4 and 1.5, respectively (Fig. 4c). Here, only the shell undergoes fragmentation, while both the polymer inner core and the cladding are cold-drawn. After dissolving the polymer, hollow glass



**Figure 4 | Characteristics of cold-drawing-driven fragmentation.**

**a**, Scalability of the in-fibre process. SEM micrograph of the fragmented rods retrieved from a 1-mm-diameter PES fibre containing about 4,000 500-nm-diameter  $\text{As}_2\text{Se}_3$  cores after cold-drawing. The left inset is a schematic of the process and the right inset highlights the size of the rods. **b**, Optical transmission micrographs before and after cold-drawing of a fibre whose core has a core-shell structure, with the inner core and cladding both PES ('P') and the shell  $\text{As}_2\text{Se}_3$  ('G'). Only the shell undergoes fragmentation, while the inner core and cladding are both cold-drawn. **c**, SEM micrographs of single hollow rods retrieved from a fibre such as in **b** with different ratios of inner (polymer-core;  $D_i$ ) to outer (glass-shell;  $D_o$ ) diameter. From left to right, parallelepiped Janus rod, hollow cylindrical Janus rod, and triangular rod with square hole. Scale bars in **c** and **d**, 10 µm. Schematics below the SEM micrographs in **c** and **d** show cross-sections of the structures, with  $\text{As}_2\text{S}_3$  ( $G_1$ ) coloured orange and  $(\text{As}_2\text{Se}_3)_{99}\text{Ge}_1$  ( $G_2$ ) coloured black. **e**, Self-healing of glass fragments after thermal restoration. **f**, Photograph showing a polymer film (PES, 'P') with gold (Au) deposited on it before and after cold-drawing. SEM micrographs (left and right columns show two different magnifications) of the fragmented 20-nm-thick and 70-nm-thick gold films.

microtubes are retrieved. Third, the geometry of the micro- or nanoparticles produced is limited only by the ability to structure the continuous core. Examples in Fig. 4d include bicompartamental parallelepiped Janus particles, hollow cylindrical Janus particles, and triangular particles with square holes. Such structures are produced

at the macroscopic scale via extrusion and then thermally drawn into a fibre with the desired transverse size; see Supplementary Information for further structures (including solid cylindrical Janus and core-shell particles). Cold-drawing then fragments the structures into particles while maintaining the complex cross-section. Fourth, the fragmentation is thermoreversible (Fig. 4e): heating the drawn fibre above its glass transition temperature results in self-healing of the fragmented core as the initial fibre dimensions are restored. Upon heating the fibre to its softening temperature, the tensile stress is released and the polymer fibre contracts to its initial length via an imperfect shape-memory effect (SME). As voids between the fractured glass segments are eliminated, the segments fuse together (self-heal) and re-form an intact longitudinal core. This effect is not a SME because—by definition—SME can occur only in samples with continuous strain histories.

Further, thermal drawing is not a necessary precursor to producing structures that undergo in-fibre fragmentation. Commercially available polymer films may be directly exploited by coating them with a relatively brittle material. For example, we show in Fig. 4f the fragmentation of a 70-nm-thick gold layer that was sputtered onto a 75- $\mu\text{m}$ -thick PES film. Cold-drawing of this composite results in fragmentation of the gold film into segments with widths of roughly 1.1  $\mu\text{m}$  along the cold-drawing axis (see Extended Data Fig. 4 for the fragmentation of gold films with thicknesses of 20–70 nm). These segments produce a clear optical-diffraction-grating signature that is visually apparent to the naked eye and indicate a grating of period of approximately 2.2  $\mu\text{m}$ , which is consistent with Fig. 4f (see Extended Data Fig. 5 for optical measurements). Alternatively, a dry-erase marker pen can be used to write straight, thick lines (widths of 2–5 mm) on the PES film (Extended Data Fig. 6). The ink interacts with the top layer of the polymer film to form an  $\sim 1\text{-}\mu\text{m}$ -thick brittle crust that fragments upon cold-drawing in a manner similar to that of the gold film.

Cold-drawing has been a mainstay of mass-production of continuous threads in the synthetic fibre and textile industries over most of the past century. It is a fundamental mechanical phenomenon that extends to the processing of macroscale metals<sup>26,27</sup> and nanocrystals<sup>28</sup>. We have shown that this process can be exploited to fabricate discontinuous three-dimensional arrays of multimaterial micro- and nanostructures. Our results suggest potential applications of cold-drawing in controlling the optical properties of macroscopic composite structures through dynamical and thermoreversible nanoscale mechanical processes. This may lead to dynamical camouflaging via a nanoscale Venetian-blind effect, scalable fabrication of micro- and nanoparticles with arbitrary cross-sections, and large-area meta-surfaces for highly sensitive detection of pathogens.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 28 May 2015; accepted 23 March 2016.**

**Published online 6 June 2016.**

- Carothers, W. H. & Hill, J. W. Studies of polymerization and ring formation. XV. Artificial fibers from synthetic linear condensation superpolymers. *J. Am. Chem. Soc.* **54**, 1579–1587 (1932).
- Vincent, P. I. The necking and cold-drawing of rigid plastics. *Polymer* **1**, 7–19 (1960).
- Coleman, B. D. On the cold drawing of polymers. *Comput. Math. Appl.* **11**, 35–65 (1985).
- Argon, A. S. *The Physics of Deformation and Fracture of Polymers* Ch. 10 (Cambridge Univ. Press, 2013).
- Carothers, W. H. Synthetic fiber. US patent 2,130,948 (1938).
- Marshall, I. & Thompson, A. B. Drawing synthetic fibres. *Nature* **171**, 38–39 (1953).
- Hermes, M. *Enough for One Lifetime: Wallace Carothers, Inventor of Nylon* (Chemical Heritage Foundation, 1996).
- Ziabicki, A. *Fundamentals of Fibre Formation: The Science of Fibre Spinning and Drawing* (Wiley, 1976).
- Carraher, C. E. Jr. *Introduction to Polymer Chemistry* 3rd edn (CRC Press, 2012) edition.

- Abouraddy, A. F. *et al.* Towards multimaterial multifunctional fibres that see, hear, sense and communicate. *Nature Mater.* **6**, 336–347 (2007).
- Tao, G., Stolyarov, A. M. & Abouraddy, A. F. Multimaterial fibers. *Int. J. Appl. Glass Sci.* **3**, 349–368 (2012).
- Loos, J., Schimanski, T., Hofman, J., Peijs, T. & Lemstra, P. J. Morphological investigations of polypropylene single-fibre reinforced polypropylene model composites. *Polymer* **42**, 3827–3834 (2001).
- Friedrich, K. *et al.* Microfibrillar reinforced composites from PET/PP blends: processing, morphology and mechanical properties. *Compos. Sci. Technol.* **65**, 107–116 (2005).
- Fakirov, S., Bhattacharyya, D., Lin, R. J. T., Fuchs, C. & Friedrich, K. Contribution of coalescence to microfibril formation in polymer blends during cold drawing. *J. Macromol. Sci. B* **46**, 183–194 (2007).
- Nairn, J. A. On the use of shear-lag methods for analysis of stress-transfer in unidirectional composites. *Mech. Mater.* **26**, 63–80 (1997).
- Asloun, E. M., Nardin, M. & Schultz, J. Stress transfer in single-fibre composites: effect of adhesion, elastic modulus of fibre and matrix, and polymer chain mobility. *J. Mater. Sci.* **24**, 1835–1844 (1989).
- Wang, X., Zhang, B., Du, S., Wu, Y. & Sun, X. Numerical simulation of the fiber fragmentation process in single-fiber composites. *Mater. Des.* **31**, 2464–2470 (2010).
- Ballato, J. *et al.* Silicon optical fiber. *Opt. Express* **16**, 18675–18683 (2008).
- Tao, G. *et al.* Infrared fibers. *Adv. Opt. Photon.* **7**, 379–458 (2015).
- Cox, H. L. The elasticity and strength of paper on other fibrous materials. *Br. J. Appl. Phys.* **3**, 72–79 (1952).
- Boyce, M. C., Montagut, E. L. & Argon, A. S. The effects of thermomechanical coupling on the cold drawing process of glassy polymers. *Polym. Eng. Sci.* **32**, 1073–1085 (1992).
- Lu, N., Suo, Z. & Vlassak, J. J. The effect of film thickness on the failure strain of polymer-supported metal films. *Acta Mater.* **58**, 1679–1687 (2010).
- Saleh, B. E. A. & Teich, M. C. *Fundamentals of Photonics* 2nd edn (Wiley, 2007).
- Zartman, G. D. *et al.* How melt-stretching affects mechanical behavior of polymer glasses. *Macromolecules* **45**, 6719–6732 (2012).
- Kaufman, J. J. *et al.* Thermal drawing of high-density macroscopic arrays of well-ordered sub-5-nm-diameter nanowires. *Nano Lett.* **11**, 4768–4773 (2011).
- Stören, S. & Rice, J. R. Localized necking in thin sheets. *J. Mech. Phys. Solids* **23**, 421–441 (1975).
- Bai, Y. & Wierzbicki, T. Forming severity concept for predicting sheet necking under complex loading histories. *Int. J. Mech. Sci.* **50**, 1012–1022 (2008).
- Fang, T. H., Li, W. L., Tao, N. R. & Lu, K. Revealing extraordinary intrinsic tensile plasticity in gradient nano-grained copper. *Science* **331**, 1587–1590 (2011).
- Schulson, E. M. The structure and mechanical behavior of ice. *J. Miner. Met. Mater. Soc.* **51**, 21–27 (1999).
- Li, T. (ed.). *Optical Fiber Communications: Fiber Fabrication* Vol. 1 (Academic Press, 1985).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank E.-H. Banaei, L. N. Pye, F. Tan, A. Schülzgen, C. Jollivet, C. Cariker, A. Schulte, M. Lodge, M. Ishigami, E. Duenas, C. Burchett, M. Finke, Y. Xu, S. Dai, H. Ren and X. Wang for technical assistance. We also thank M. Rein, F. Sorin, M. Kolle, A. Dogariu, D. N. Christodoulides and B. E. A. Saleh for discussions. The authors acknowledge the University of Central Florida Stokes Advanced Research Computing Center for providing computational resources and support that have contributed to results reported here. We also thank Simulia, Inc. for providing the license of the ABAQUS software package. This work was supported by the US Air Force Office of Scientific Research (AFOSR) under contract FA-9550-12-1-0148 and AFOSR MURI contract FA9550-14-1-0037, and the US National Science Foundation (CMMI-1300773). This work was supported in part by the MIT MRSEC through the MRSEC Program of the National Science Foundation under award number DMR-1419807.

**Author Contributions** S.S. and A.F.A. developed and directed the project. S.S. observed the cold-drawing-related phenomena, carried out the break-up experiments in fibres and films, produced the hybrid samples based on hollow-core polymer fibres, and performed the optical measurements, the particle length measurements and the thermal restoration experiments. G.T. extruded the multimaterial preforms, fabricated the stack-and-draw preforms and produced all the chalcogenide-polymer and tellurite-polymer fibres. J.J.K. carried out the SEM imaging and produced the hollow-core polymer fibres and the multi-core fibres. Y.Q. and Y.B. carried out the finite-element simulations. L.W. and Y.F. produced the thin glass films and the Si and Ge micro-wires. T.B., S.S. and A.P.G. performed the stress-strain measurements and recorded the Supplementary Videos. R.S.H. developed the heuristic analytical model. Y.F., Y.B., R.S.H. and A.F.A. supervised the research. S.S., Y.B., R.S.H. and A.F.A. wrote the paper. All authors contributed to the interpretation of the results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.F.A. ([raddy@creol.ucf.edu](mailto:raddy@creol.ucf.edu)).

## METHODS

**Fibre fabrication.** The polymer cylindrical (Figs 1 and 2) and flat (Fig. 3) PES fibres are produced by thermal drawing in a custom-built fibre draw tower. In case of core materials that are thermally compatible with PES (that is, they soften in an overlapping temperature range in which they have comparable viscosity<sup>10,11</sup>), such as chalcogenide and tellurite glasses, the polymer cladding and glass core are co-drawn into a fibre from a centimetre-scale 'preform'. The fibre pre-stress is set during the thermal drawing process; thermal drawing is followed by a rapid quench. In the case of all other materials that are not thermally compatible with PES, microfibrils of these materials are first prepared (see Supplementary Information) and then placed within a hollow-core PES fibre, which is thermally drawn from a hollow preform. The fibre assembly is heated to enable collapse of the PES around the core material and thus to produce strong adhesion between PES and the core.

**Thin-film preparation.** The glassy chalcogenide films (Fig. 3) used are thermally evaporated under vacuum onto a thin PES film that is then incorporated into a flat preform, thermally consolidated under vacuum, and subsequently drawn into a fibre. The gold films (Fig. 4f) are deposited using a sputtering system (Cressington 108) onto commercially available 75- $\mu\text{m}$ -thick PES thin films (Ajedum). The thickness of the gold is controllable up to 50 nm by varying the sputter time. Multiple sputtering procedures are performed to achieve thicknesses exceeding 50 nm.

**Stress-strain measurements.** Stress-strain measurements of the fibres are gathered from uniaxial tension testing performed via an Instron universal test machine with a load cell resolution of  $\pm 0.01$  N and the specimen ends are fixed using TestResources wave-type grips. The initial clear length of the fibre between the two grips is typically at least 20 cm. Each test is conducted under displacement-control of  $10\text{ mm min}^{-1}$  to rupture.

**Simulations for cylindrical fibres.** The Explicit solver of the general-purpose finite-element code ABAQUS<sup>31</sup> is used for the computational analysis of the cold-drawing and fragmentation processes. For the cylindrical fibres (Fig. 2e), an axisymmetric model is built using CAX4R elements<sup>31</sup> ( $5\text{ }\mu\text{m} \times 5\text{ }\mu\text{m}$ ) consisting of three concentric 2.52-mm-long cylindrical layers (Extended Data Fig. 2a). The inner section (10- $\mu\text{m}$  radius) corresponds to the core material, the cladding section (500- $\mu\text{m}$  radius) to the PES matrix. The stress-strain curve for PES (Extended Data Fig. 2b) is based on the test data (Fig. 2c). These two layers are separated by an interfacial layer<sup>32</sup> that possesses similar properties to PES, but with much weaker strength; we thus take the stress-strain curve for this interfacial layer to be the same as for PES, but multiplied by a scalar factor that depends on the core material. Extended Data Fig. 2c shows the stress-strain curve for the interfacial layer when the core is  $\text{As}_2\text{Se}_3$ . For the core section, we use a brittle-cracking failure model<sup>33</sup> available in ABAQUS, which includes a linear elastic range (taken from test data; see Supplementary Fig. 14f) and softening behaviour at crack propagation. We use an explicit code to predict necking in the PES cladding<sup>34–36</sup>. Specifically, we use the elastic-viscoplastic continuum model proposed in ref. 36. The model is calibrated with respect to the test data and implemented in ABAQUS using the material subroutine VUMAT. In the simulations, the lower grip is fixed and the upper grip is assigned a constant velocity of  $5\text{ mm s}^{-1}$  upwards. Simulations are conducted for multiple core materials, including  $\text{As}_2\text{Se}_3$  (Extended Data Fig. 2d), Si (Extended Data Fig. 2e), Ge and polystyrene. The results are summarized in Extended Data Table 2. The average length-to-diameter ratios  $\bar{L}/D$  of fragmented rods and their standard deviations correlate well with the measurements (Fig. 2b).

**Simulations for flat fibres.** The flat-fibre cold-drawing simulations are carried out using a similar procedure to that for the cylindrical-fibre simulations. A plane-strain model is built in the plane spanned by the longitudinal axis of the fibre and the fibre thickness ( $y$ - $z$  plane in Fig. 3a). This is a good approximation for the thin film because the strain in the transverse direction is much smaller than that in the other two directions during cold-drawing. A quarter of the cross-section is modelled, with symmetric boundary conditions, using CPE4R elements<sup>31</sup> approximately  $75\text{ nm} \times 1,000\text{ nm}$  in size. There are three sections defined in the simulation: the film, an interfacial layer and the outer PES cladding. The dimensions used in the simulation (Extended Data Fig. 3) are the same as those in the experiment (Fig. 3a): the film and fibre thicknesses are 300 nm and 350  $\mu\text{m}$ , respectively, and the initial fibre length is 1.44 mm. The material models used are the same as those used in the cylindrical cold-drawing simulations above (Extended Data Fig. 2). The moving grip applies a constant tensile velocity of  $2\text{ mm s}^{-1}$ . Extrusion and mirror methods<sup>31</sup> are used to visualize the results in three dimensions. The width of the fragmented strips in the cold-drawing simulation is  $7.55\text{ }\mu\text{m}$ , which is in good agreement with the measurements.

**Animations.** Supplementary Videos 2 and 4 show simulations of the cylindrical fibre and of the flat fibre, respectively. In each video, the frames of the above-described simulations are assembled, in addition to a view of the insets shown in

Fig. 2e (for the cylindrical fibre) and Extended Data Fig. 3 (for the flat fibre). In both videos, the speed of the frames is reduced by a factor of 100 compared to the time steps of the simulation drawing speed.

**Optical model for transmission through fragmented glass films.** We consider an optical model of the flat fibre in Fig. 3a consisting of a 300-nm-thick film of  $\text{As}_2\text{Se}_3$  embedded in a 300- $\mu\text{m}$ -thick, 2-mm-wide rectangular PES fibre (Fig. 3a). The refractive index of PES  $n_p(\lambda)$  is extracted from spectroscopic ellipsometry measurements in the wavelength range  $\lambda = 0.35\text{--}2\text{ }\mu\text{m}$ . The refractive index of the glass  $n_g(\lambda)$  is obtained from a Sellmeier equation<sup>37</sup> and experimentally confirmed by ellipsometric measurements, while spectral absorption in the visible range is obtained from optical transmission measurements; optical transmission through the thin glass film is negligible when  $\lambda < 0.5\text{ }\mu\text{m}$ . The glass-polymer interfaces give rise to optical-field Fresnel reflection and transmission coefficients  $r_{pg}(\lambda)$  and  $t_{pg}(\lambda)$ ; the two air-polymer interfaces contribute a transmission coefficient  $t_p(\lambda)$ . Transmission through the intact flat fibre is  $T_{\text{intact}} = |t_p|^2 |t_{pg}|^2 [1 + r_{pg}^2 e^{i\frac{4\pi}{\lambda} n_g d}]^2$ , in which  $d$  is the thickness of the glass film and we have taken into consideration only the first two Airy wavelets in the lossy Fabry-Perot resonator formed of the glass film (absorption in the film diminishes the effect of higher-order reflected wavelets). Transmission through a cold-drawn sample embedding the fragmented glass strips is calculated by weighing the transmission through an intact sample  $T_{\text{intact}}$  and the transmission through a polymer sample  $T_p = |t_p|^2$  with respect to the relative areas of the fragmented strips and the voids separating them. If we define  $\delta$  to be the ratio of the width of a void separating two strips to the width of a strip, then the transmission through the cold-drawn sample is  $T = (\alpha T_p + \beta T_{\text{intact}}) \eta_{\text{diff}}$ , in which  $\alpha = \frac{1}{1+\delta}$ ,  $\beta = 1 - \alpha$ ,  $\eta_{\text{diff}}(\lambda) = \frac{\int_{\text{det}} I(x) dx}{\int_{\text{total}} I(x) dx}$  is the fraction of light diffracted

through the evenly spaced glass strips that reaches the detector aperture as function of wavelength, and  $I(x)$  is the intensity distribution in the detector plane. Using the measured values of the refractive indices of the fibre materials in these formulae, we calculate the theoretical optical transmission spectra plotted in Fig. 3e.

**Diffraction measurements on fragmented gold films.** The gold-PES cold-drawn samples diffract transmitted (and reflected, Fig. 4f) white light, leading to angular-selective visible colours. To quantify this observation, the diffracted spectra are measured in transmission mode as a function of the angle with respect to the axis defined by normal incidence. A broadband, incoherent, white-light optical source is coupled into a multimode optical fibre by means of a fibre collimator, and out-coupled light is collimated using a  $\times 4$  microscope objective lens. The  $5\text{ mm} \times 20\text{ mm}$  sample is placed at a distance of 10 cm from the lens, whereupon the beam diameter is approximately 4 mm. Diffracted light at angle  $\theta$  is coupled into a multimode fibre via a lens with a 50-mm focal length, and the spectra are recorded using an optical spectrum analyser (OSA; Advantest Q8381A); see Extended Data Fig. 5a. The diffracted light is blueshifted at small angles and redshifted at large angles (Extended Data Fig. 5b).

Assuming normal incidence on the film, the ideal grating equation indicates that  $\lambda = \Lambda \sin(\theta)$ , in which  $\lambda$  is the wavelength,  $\Lambda$  is the grating period and  $\theta$  is the diffraction angle of the first diffraction order with respect to the normal to the film. We identify  $\theta$  with the peak diffracted wavelength  $\lambda_{\text{max}}$  in Extended Data Fig. 5b, so that we have the following pairings ( $\lambda_{\text{max}}$ ,  $\theta$ ): (420 nm,  $10.6^\circ$ ), (700 nm,  $17.4^\circ$ ) and (800 nm,  $21.2^\circ$ ). These measurements indicate grating periods  $\Lambda$  of 2.3  $\mu\text{m}$ , 2.34  $\mu\text{m}$  and 2.21  $\mu\text{m}$ , respectively, which are consistent with the SEM micrographs in Fig. 4f.

**Preparing ink-written films for cold-drawing.** The ink-PES samples are prepared by cutting a 75- $\mu\text{m}$ -thick PES film (Ajedum) into strips with dimensions of about  $5\text{ mm} \times 10\text{ cm}$ . Using a black permanent marker, thick lines (approximately 2–5-mm thick) are painted on one face of the strips. The marker tip is traced on the PES film only once to avoid non-uniformities in the ink layer left on the film surface (Extended Data Fig. 6a). The samples are dried for a few seconds and then cold-drawn by symmetrically pulling both ends using a pair of pliers (Extended Data Fig. 6b). The optical appearance of the film (Extended Data Fig. 6c) changes immediately after cold-drawing, whereupon bright coloured optical diffraction bands are visible to the eye. SEM micrographs of the surface of the film (Extended Data Fig. 6d, e) reveal that the ink absorbed at the surface of the film produces a crust with an average thickness of  $< 1\text{ }\mu\text{m}$ . The thickness of the crust tapers at the edges of the drawn line. We find that the ink crust after cold-drawing fragments into strips whose width is proportional to the crust thickness. The ink-crust strips are parallel and aligned orthogonally to the cold-drawing axis (the long dimension of the polymer strip), similarly to the observations in Fig. 3 with the glass films embedded in a flat polymer fibre. Thus, at the edge of the line where the thickness drops rapidly, we concomitantly observe a rapid drop in the width of the strips (Extended Data Fig. 6e).

**Cold-drawing break-up versus stress transfer in fibre-reinforced composites.** The results presented here are related to the shear-lag fracture (SLF) phenomena

often observed in fibre-reinforced composites. However, the fragmentation we report is distinct from canonical SLF in both aim and character, as follows.

(1) In traditional composites, the role of the core is to strengthen the fibre. By contrast, our goal is to exploit the cold-drawing process to produce controlled fragmentation of the core. In other words, we exploit SLF rather than seeking to avoid it as has been the focus of most previous work.

(2) There are several aspects that distinguish the SLF process reported here from canonical SLF. Since the seminal work by Cox<sup>20</sup>, the standard theoretical frameworks for analysing stress transfer in composites have used uniaxial models. Most treatments have assumed that there is no load transfer at the fibre ends because they are both 'free'. By contrast, our fibres are semi-infinite; fracture occurs at the free end, while the other end remains tethered to the remainder of the core. More fundamentally, cold-drawing necessarily produces non-trivial multi-axial stresses<sup>4,21</sup>, especially in thin films<sup>22</sup>. We expect that these distinguishing features will stimulate future theoretical and experimental investigations into SLF.

(3) In contrast to canonical SLF where fragmentation typically occurs simultaneously in many locations, our core fibres fragment only within the advancing neck front. Once a fragment of the core is separated, it does not undergo any further fracture.

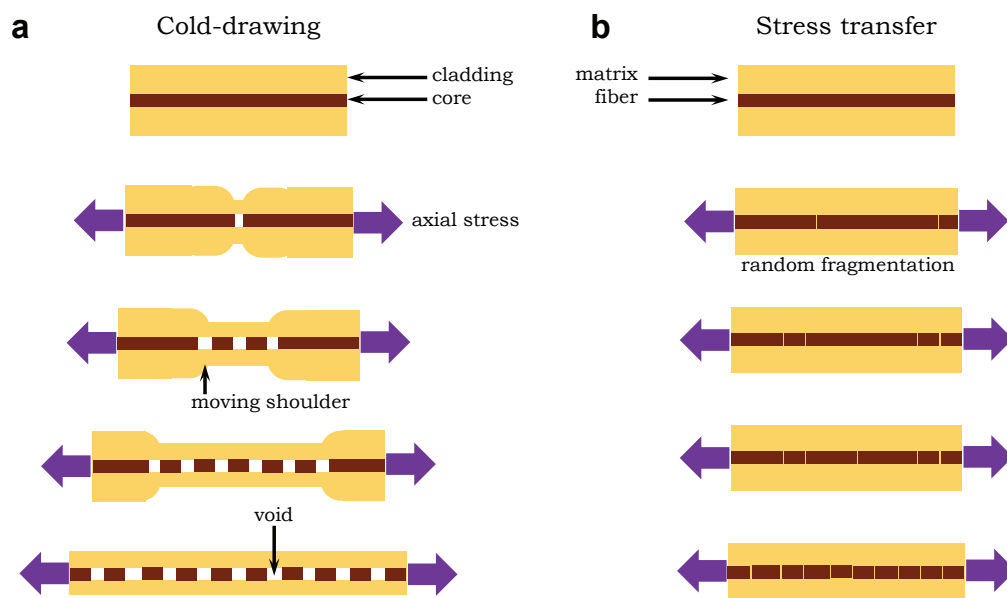
(4) Core fracture occurs as the core passes through the propagating shoulder, producing multimaterial rods of controllable length. Alternatively, when one of the materials making up the core is the same polymer used as the cladding, this portion of the core/fibre undergoes cold-drawing and remains intact while the remaining materials fracture. This feature enables the production of hollow structures via selective dissolution of polymer.

Extended Data Fig. 1 illustrates the distinction between SLF that takes place in the process reported here and in tradition fibre-reinforced composites. The nature of each fracture event taking place during the necking of the polymer fibres is the same as that occurring in SLF in composite materials. In both cases, a local brittle fracture takes place. The fundamental difference lies in the global dynamics of the break-up process.

**Traditional SLF.** In the case of traditional SLF<sup>38–41</sup> (Extended Data Fig. 1b), the setting is typically that of a composite material, usually fibres in a matrix (what we call core in a cladding). When axial stress is applied, random brittle fractures occur along the fibres. This process continues as long as the stress is applied or until a minimum size of the fractured fibre segments is reached. If the stress is removed before saturation is reached, then what remains is a collection of fibre fragments of unequal, random lengths. Re-applying the axial stress leads to a continuation of the brittle fracture (still localized in the shoulder) until saturation. Post-saturation, the axial stress does not lead to further fracture.

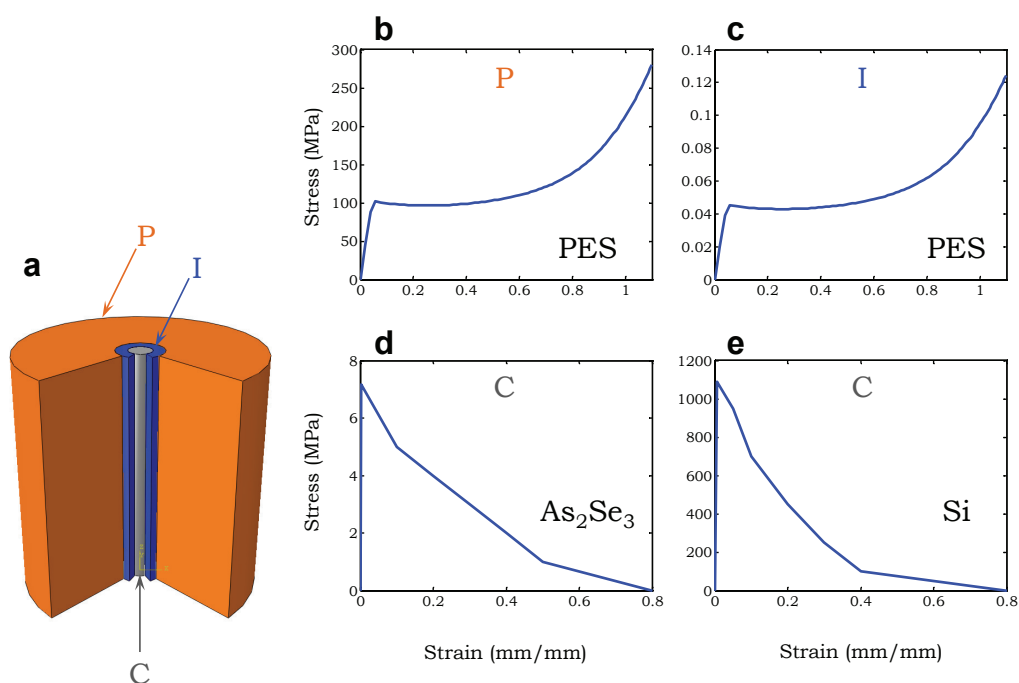
**Fragmentation during cold-drawing.** In the case of fragmentation during cold-drawing (Extended Data Fig. 1a), the polymer cladding (or matrix) undergoes necking that is by and large independent of the brittle core. During the propagation of the neck upon axial stress, the core undergoes brittle fracture in an orderly fashion on a global scale: the fracture occurs sequentially within the moving shoulder. This process continues as long as the axial stress is applied or until the fibre becomes fully drawn. If the axial stress is removed before the fibre is fully drawn, then we are left with a section of the core that is intact and that has not undergone fracture (the section with the undrawn polymer), and a section that has fractured in a periodic sequence of equally sized fragments (the drawn section of the polymer fibre). Re-applying the axial stress results in a continuation of necking and resumption of the associated fragmentation. After the fibre is fully drawn, no more fragmentation takes place.

31. ABAQUS User's Manual, version 6.11 (Simulia, Inc., 2011).
32. Mital, S. K., Murthy, P. L. N. & Chamis, C. C. Interfacial microfracture in high temperature metal matrix composites. *J. Compos. Mater.* **27**, 1678–1694 (1993).
33. Hillerborg, A., Mod  r, M. & Petersson, P.-E. Analysis of crack formation and crack growth in concrete by means of fracture mechanics and finite elements. *Cement Concr. Res.* **6**, 773–781 (1976).
34. Hutchinson, J. W. & Neale, K. W. Neck propagation. *J. Mech. Phys. Solids* **31**, 405–426 (1983).
35. Neale, K. W. & Tugcu, P. Analysis of necking and neck propagation in polymeric materials. *J. Mech. Phys. Solids* **33**, 323–337 (1985).
36. Anand, L. & Gurtin, M. E. A theory of amorphous solids undergoing large deformations, with application to polymeric glasses. *Int. J. Solids Struct.* **40**, 1465–1487 (2003).
37. Ung, B. & Skorobogatiy, M. Chalcogenide microporous fibers for linear and nonlinear applications in the mid-infrared. *Opt. Express* **18**, 8647–8659 (2010).
38. Figueroa, J. C., Carney, T. E., Schadler, L. S. & Laird, C. Micromechanics of single filament composites. *Compos. Sci. Technol.* **42**, 77–101 (1991).
39. Galiotis, G. & Paipetis, A. Definition and measurement of the shear-lag parameter,  $\beta$ , as an index of the stress transfer efficiency in polymer composites. *J. Mater. Sci.* **33**, 1137–1143 (1998).
40. Kim, B. W. & Nairn, J. A. Observations of fiber fracture and interfacial debonding phenomena using the fragmentation test in single fiber composites. *J. Compos. Mater.* **36**, 1825–1858 (2002).
41. Thostenson, E. T., Li, W. Z., Wang, D. Z., Ren, Z. F. & Chou, T. W. Carbon nanotube/carbon fiber hybrid multiscale composites. *J. Appl. Phys.* **91**, 6034–6037 (2002).
42. Kurkjian, C. R. Mechanical properties of phosphate glasses. *J. Non-Cryst. Solids* **263–264**, 207–212 (2000).
43. Chang, K. H., Lee, T. H. & Hwa, L. G. Structure and elastic properties of iron phosphate glasses. *Chin. J. Physiol.* **41**, 414–421 (2003).
44. Rouxel, T. Elastic properties and short- to medium-range order in glasses. *J. Am. Ceram. Soc.* **90**, 3019–3039 (2007).
45. P  rez-Rigueiro, J., Viney, C., Llorca, J. & Elices, M. Mechanical properties of single-brin silkworm silk. *J. Appl. Polym. Sci.* **75**, 1270–1277 (2000).
46. Cheung, H.-Y., Lau, K.-T., Ho, M.-P. & Mosallam, A. Study on the mechanical properties of different silkworm silk fibers. *J. Compos. Mater.* **43**, 2521–2531 (2009).
47. Zhang, K., Si, F. W., Duan, H. L. & Wang, J. Microstructures and mechanical properties of silks of silkworm and honeybee. *Acta Biomater.* **6**, 2165–2171 (2010).
48. Hartouni, E. & Mecholsky, J. J. Mechanical properties of chalcogenide glasses. *Proc. SPIE* **0683**, 92–97 (1986).
49. Littler, I. C. M., Fu, L. B., M  gi, E. C., Pudo, D. & Eggleton, B. J. Widely tunable, acousto-optic resonances in chalcogenide As<sub>2</sub>Se<sub>3</sub> fiber. *Opt. Express* **14**, 8088–8095 (2006).
50. Tanaka, K. & Shimakawa, K. *Amorphous Chalcogenide Semiconductors and Related Materials* (Springer, 2011).
51. El-Mallawany, R. A. H. *Tellurite Glasses Handbook: Physical Properties and Data* (CRC Press, 2011).
52. Hu, Z. *et al.* Measurement of Young's modulus and Poisson's ratio of human hair using optical techniques. *Proc. SPIE* **7522**, 75222Q (2010).
53. Kaplan, P. D. *et al.* Grey hair: clinical investigation into changes in hair fibres with loss of pigmentation in a photoprotected population. *Int. J. Cosmet. Sci.* **33**, 171–182 (2011).
54. Nie, H.-Y., Motomatsu, M., Mizutani, W. & Tokumoto, H. Local modification of elastic properties of polystyrene-polyethyleneoxide blend surfaces. *J. Vac. Sci. Technol. B* **13**, 1163–1166 (1995).
55. Bellan, L. M., Kameoka, J. & Craighead, H. G. Measurement of the Young's moduli of individual polyethylene oxide and glass nanofibers. *Nanotechnology* **16**, 1095–1099 (2005).
56. Weeks, W. F. & Assur, A. *The Mechanical Properties of Sea Ice* (US Army Material Command, Cold Regions Research and Engineering Laboratory, 1967).



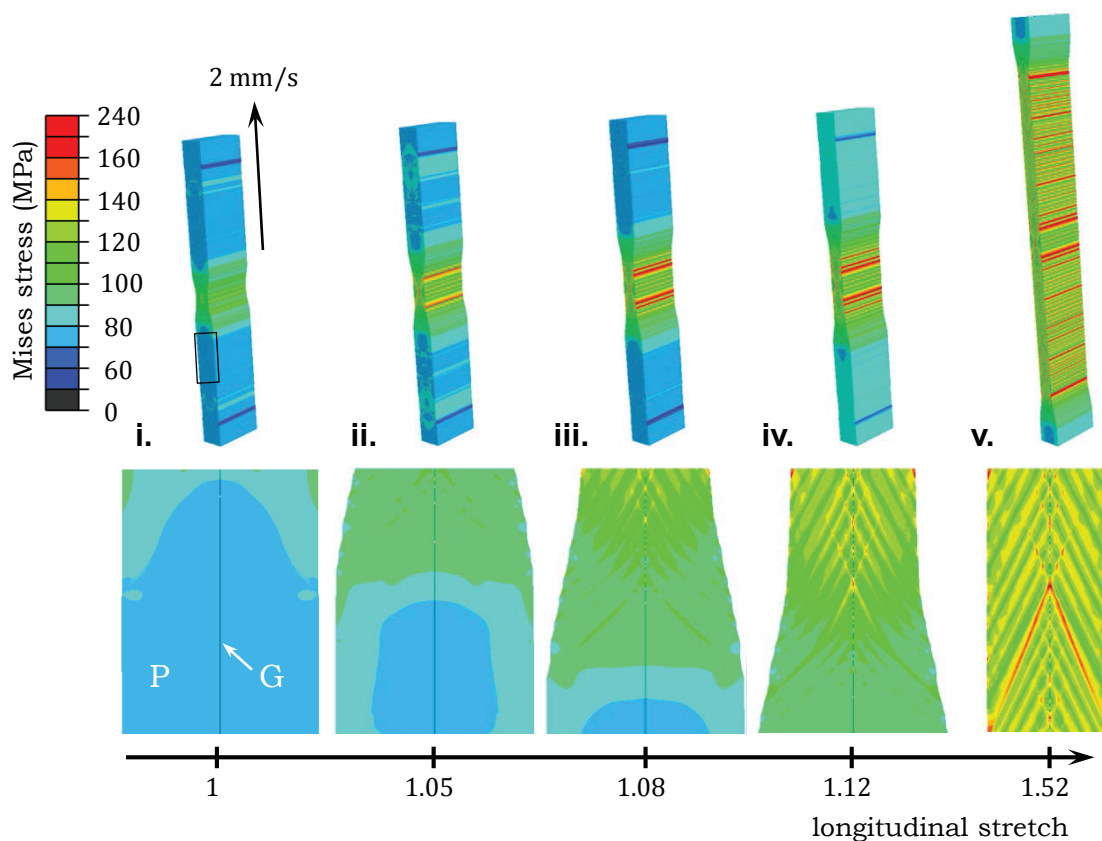
**Extended Data Figure 1 | Schematic contrasting controlled (sequential) and uncontrolled (random) thread fragmentation.** **a**, The designed fragmentation process that takes place during cold-drawing of a fibre consisting of a brittle core embedded in a ductile cladding. The overall

length of the sample increases considerably when fully drawn. **b**, The random fragmentation that takes place during stress transfer in a composite sample consisting of a fibre embedded in a matrix. Thick purple arrows indicate externally applied stress.



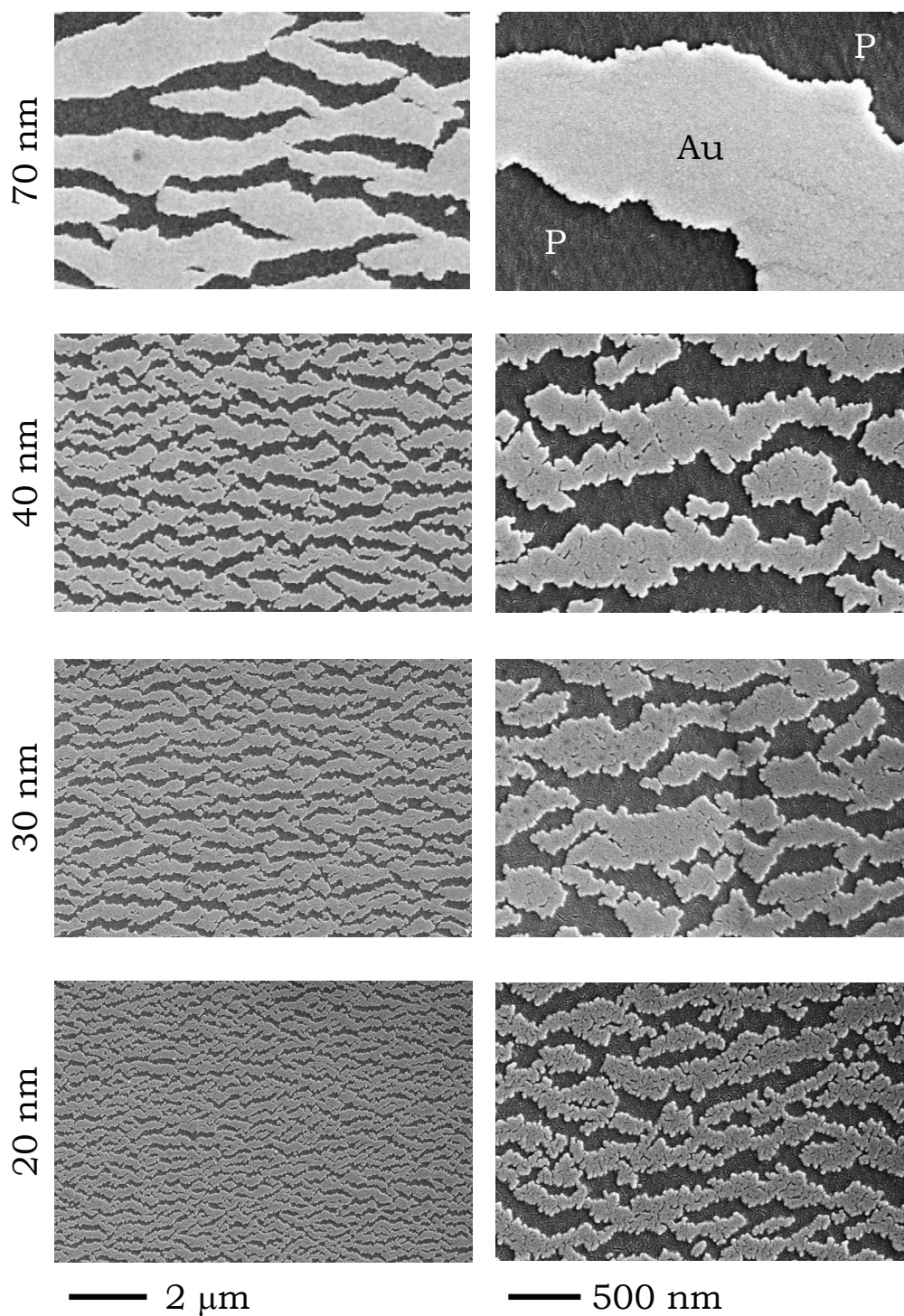
**Extended Data Figure 2 | Stress-strain models of the materials used in the finite-element computational model.** **a**, Axisymmetric structure used in the computational model. P, polymer (PES); I, interfacial layer; C, core. The same polymer and interfacial layer are used in the cylindrical and flat fibre simulations. Various core materials are used. **b**, Stress-strain model

for the PES cladding ('P' in **a**). **c**, Stress-strain model (including both elastic range and post-failure softening) for the PES interfacial layer ('I' in **a**). **d**, Stress-strain model for an  $\text{As}_2\text{Se}_3$  core material ('C' in **a**). **e**, Stress-strain model for a silicon (Si) core; see Supplementary Figs 14 and 15.



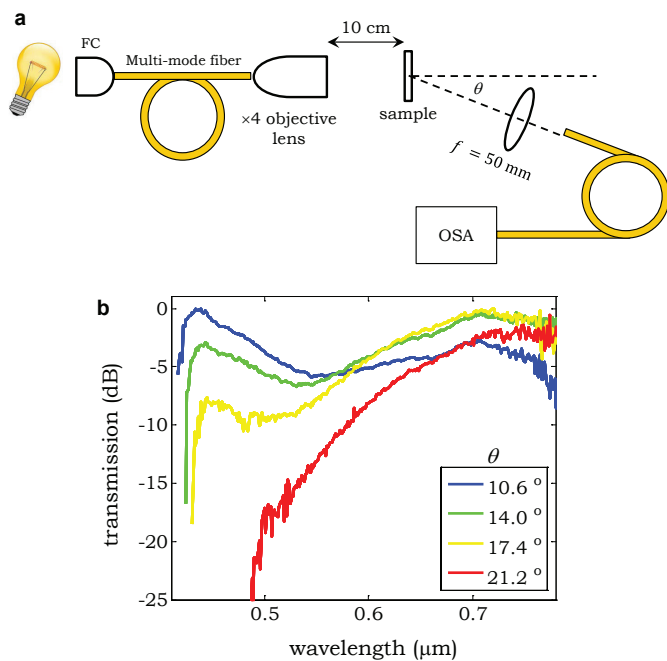
**Extended Data Figure 3 | Simulation of cold-drawing in a flat fibre containing a thin brittle film.** The results of nonlinear finite-element simulations showing contour plots of the evolving von Mises stress distribution with increasing stretch, using the same (isotropic) materials (PES and  $\text{As}_2\text{Se}_3$ ) as in the cylindrical case (Fig. 2e and Extended Data

Fig. 2). The five steps (i)–(v) correspond to increasing stretch values. Top panels depict the full fibre; bottom panels show the regions corresponding to that highlighted by the rectangle in (i). P, polymer (PES); G, glass ( $\text{As}_2\text{Se}_3$ ).

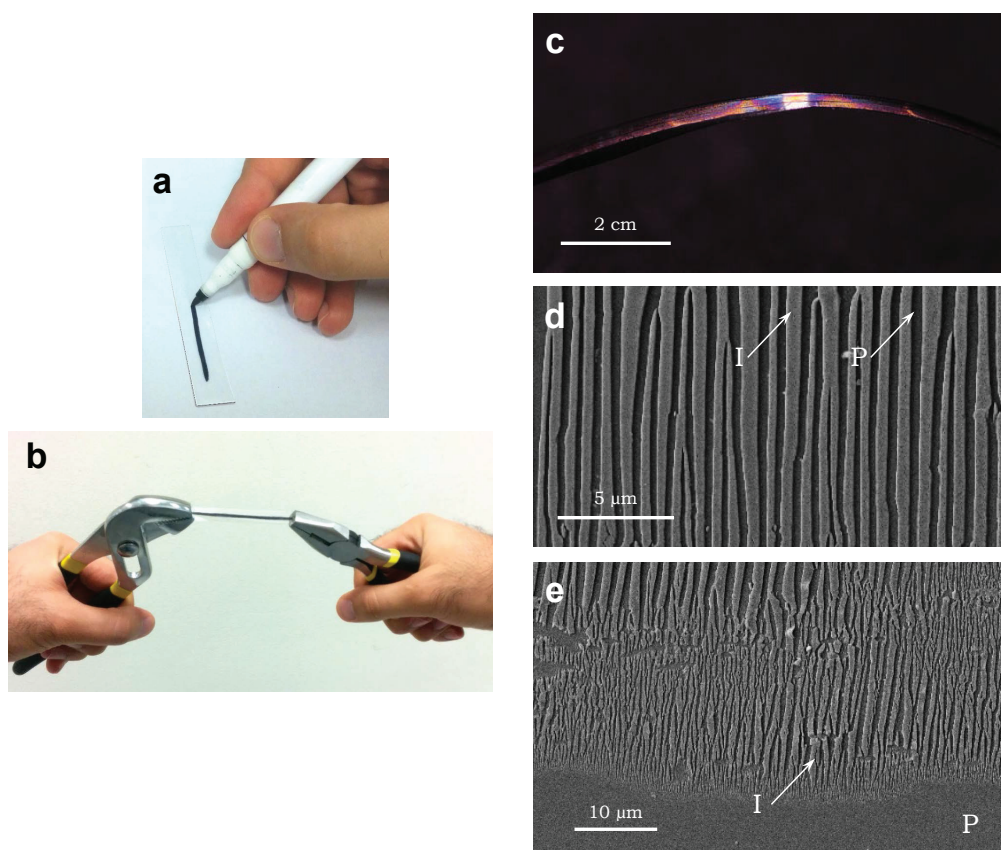


**Extended Data Figure 4 | Fragmentation of a gold thin film under cold-drawing.** Each row corresponds to a different thickness of gold (20 nm, 30 nm, 40 nm and 70 nm) sputtered onto a 75- $\mu\text{m}$ -thick PES film. The

columns show SEM micrographs of the gold films after cold-drawing at two different scales to highlight the dependence of the average fragment size on the thickness of the gold layer. P, PES film; Au, sputtered gold.



**Extended Data Figure 5 | Spectral diffraction measurements from a gold film fragmented by cold-drawing.** **a**, Optical set-up used to measure the spectrum of light diffracted at an angle  $\theta$  from a thin gold film of thickness 70 nm on a 75- $\mu\text{m}$ -thick PES film after fragmentation via cold-drawing (Extended Data Fig. 4, first row). OSA, optical spectrum analyser; FC, fibre coupler;  $\theta$  is the angle with respect to normal incidence on the film. **b**, Measured diffracted spectra on a vertical logarithmic scale. The spectra are normalized with respect to the input optical spectrum. Each spectrum is then normalized to its maximum value.



**Extended Data Figure 6 | Fragmentation of a layer of ink on PES under cold-drawing.** **a, b,** Photographs depicting the cold-drawing procedure. **a,** A line is drawn on a 75- $\mu\text{m}$ -thick PES film (5 mm  $\times$  10 cm) using a dry-erase marker pen. **b,** Using two pliers, the two ends of the strip are pulled symmetrically by hand until cold-drawing is complete. **c,** After cold-drawing, the optical appearance of the strip changes and coloured diffracted bands are apparent to the naked eye (the marker pen is used to

write across the whole film surface). **d,** SEM micrograph of the drawn line reveals that a crust is formed at the PES surface that fragments into strips that are orthogonal to the cold-drawing axis (similarly to in Figs 3 and 4f), which are behind the new optical properties of the strip seen in **c**. **e,** SEM micrograph of the edge of the drawn line, showing a tapering of the thickness of the ink crust, and concomitant drop in fragmentation period. I, ink-polymer crust; P, PES polymer film.

**Extended Data Table 1 | Measured values of  $\bar{L}/D$  and measured and reported values of  $E$ , for different core materials**

Material	$\bar{L}/D$	St. Dev.	$E$ (GPa)	$\Delta$ (GPa)	$E$ (GPa)	$\Delta$ (GPa)
	Measured		Measured		Reported	
SiO <sub>2</sub>	33	11.95	88.5	5.5	70	65 – 73.1
Si	32.5	9.56	152	7.5	150	130–202
Ge	30	9.26	78	6	103	102.7–103
PhG <sup>42–44</sup>	9.5	3.19	11	2	55	31.3–79
Silk <sup>45–47</sup>	8.2	2.07	9	3.2	16	3.8–17
ChG <sup>48–50</sup>	6	1.57	3.2	0.1	38.6	18–40
TeG <sup>51</sup>	3.9	1.07	N/A <sup>(1)</sup>	N/A <sup>(1)</sup>	44	37.1–50.7
Hair <sup>52,53</sup>	4.9	1.46	3.7	1.1	4	2.5–7.5
PEO <sup>54,55</sup>	2.9	0.56	0.3	0.1	5	0.2–7
Ice <sup>20,44,56</sup>	1.7	0.46	N/A <sup>(2)</sup>	N/A <sup>(2)</sup>	5.1	0.3–10
PS	1.6	0.25	2.7	0.7	3.2	3–3.5

The measures values of  $\bar{L}/D$  are means, with the standard deviation ('St. Dev.') also given for each material. The measured and reported values of  $E$  are means, with  $E \pm \Delta$  indicating the range of measured values, and  $\Delta$  indicating the range of reported values. PhG, phosphate glass; ChG, chalcogenide glass; TeG, tellurite glass; PEO, polyethylene oxide; PS, polystyrene. Reported values are from refs 42–56, as indicated.

<sup>(1)</sup>We did not measure  $E$  for the TeG used in our experiments. We produced polymer-TeG fibres by thermal co-drawing, which requires that the polymer and the glass have overlapping softening temperatures  $T_{\text{soft}}$  (in our context, this refers to the temperature at which the viscosity drops to values compatible with thermal drawing, typically  $10^4$ – $10^6$  Poise). All TeG materials reported in the literature have a  $T_{\text{soft}}$  that is substantially higher than that of engineering thermoplastic polymers such as PES. We modified the TeG composition to reduce  $T_{\text{soft}}$ . However, the TeG composition we used is hygroscopic; while the TeG is isolated from the ambient environment within the PES cladding, it remains stable; once the PES cladding is removed (to measure  $E$ ), the TeG takes up moisture and becomes brittle. We expect that  $E$  is reduced when the  $T_{\text{soft}}$  of the material is reduced.

<sup>(2)</sup>We did not measure  $E$  for ice and instead plot in Fig. 2b the range of values that have been reported in the literature. This range is quite large, because  $E$  for ice depends on temperature, pressure, contaminants in the water, and so on. It is difficult to determine exactly the temperature of the ice during necking, but it is expected that the ice is in the process of melting and thus we estimate that  $E$  is on the lower end of the specified range.

**Extended Data Table 2 | Comparison of simulated and measured (Fig. 2b) average value of  $L/D$  (denoted  $\bar{L}/D$ ) and its standard deviation for different core materials**

Core material	Simulations		Measurements	
	$\bar{L}/D$	St. Dev. in $L/D$	$\bar{L}/D$	St. Dev. in $L/D$
As <sub>2</sub> Se <sub>3</sub>	5.6	1.4	6	1.57
Si	36.6	5.9	32.5	9.56
Ge	30.9	5.4	30.0	9.26
Polystyrene (PS)	1.72	0.78	1.6	0.25

The cladding polymer is PES.

# Abiological catalysis by artificial haem proteins containing noble metals in place of iron

Hanna M. Key<sup>1,2\*</sup>, Paweł Dydio<sup>1,2\*</sup>, Douglas S. Clark<sup>3,4</sup> & John F. Hartwig<sup>1,2</sup>

Enzymes that contain metal ions—that is, metalloenzymes—possess the reactivity of a transition metal centre and the potential of molecular evolution to modulate the reactivity and substrate-selectivity of the system<sup>1</sup>. By exploiting substrate promiscuity and protein engineering, the scope of reactions catalysed by native metalloenzymes has been expanded recently to include abiological transformations<sup>2,3</sup>. However, this strategy is limited by the inherent reactivity of metal centres in native metalloenzymes. To overcome this limitation, artificial metalloproteins have been created by incorporating complete, noble-metal complexes within proteins lacking native metal sites<sup>1,4,5</sup>. The interactions of the substrate with the protein in these systems are, however, distinct from those with the native protein because the metal complex occupies the substrate binding site. At the intersection of these approaches lies a third strategy, in which the native metal of a metalloenzyme is replaced with an abiological metal with reactivity different from that of the metal in a native protein<sup>6–8</sup>. This strategy could create artificial enzymes for abiological catalysis within the natural substrate binding site of an enzyme that can be subjected to directed evolution. Here we report the formal replacement of iron in Fe-porphyrin IX (Fe-PIX) proteins with abiological, noble metals to create enzymes that catalyse reactions not catalysed by native Fe-enzymes or other metalloenzymes<sup>9,10</sup>. In particular, we prepared modified myoglobins containing an Ir(Me) site that catalyse the functionalization of C–H bonds to form C–C bonds by carbene insertion and add carbenes to both  $\beta$ -substituted vinylarenes and unactivated aliphatic  $\alpha$ -olefins. We conducted directed evolution of the Ir(Me)-myoglobin and generated mutants that form either enantiomer of the products of C–H insertion and catalyse the enantio- and diastereoselective cyclopropanation of unactivated olefins. The presented method of preparing artificial haem proteins containing abiological metal porphyrins sets the stage for the generation of artificial enzymes from innumerable combinations of PIX-protein scaffolds and unnatural metal cofactors to catalyse a wide range of abiological transformations.

To create artificial metalloenzymes formed by combining abiological metals and natural metalloprotein scaffolds, we focused on haem proteins, which contain Fe-porphyrin IX (Fe-PIX) as a metal cofactor. Native haem enzymes catalyse reactions that include C–H oxidation and halogenation<sup>11</sup>, and they have been successfully evolved to oxidize abiological substrates<sup>12,13</sup>. Fe-PIX proteins have also been shown to catalyse abiological reactions involving the addition and insertion of carbenes and nitrenes to olefins and X–H bonds<sup>2,3,9,14</sup>. However, the reactivity of the Fe-centre in haem proteins limits the scope of these transformations. For example, Fe-PIX proteins catalyse the cyclopropanation of activated terminal vinylarenes<sup>9,10</sup>, but they do not catalyse reactions with internal vinylarenes or unactivated alkenes. Likewise, they catalyse insertions of carbenes into reactive

N–H and S–H bonds, but they do not catalyse the insertion into less reactive C–H bonds<sup>3,14</sup>.

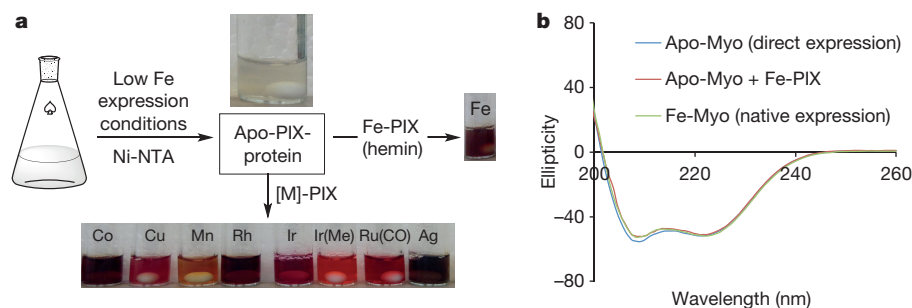
Because the repertoire of reactions catalysed by free metal-porphyrin complexes of Ru (ref. 15), Rh (ref. 16) and Ir (ref. 17) is much greater than that of the free Fe-analogues, we hypothesized that their incorporation into PIX proteins could create new enzymes for abiological catalysis that is not possible with Fe-PIX enzymes. Artificial PIX proteins containing Mn, Cr, and Co cofactors have been prepared to mimic the intrinsic chemistry of the native haem proteins<sup>18–21</sup>, but the reactivities and selectivities of these processes are lower than those achieved in the same reactions catalysed by native Fe-PIX enzymes. Thus, artificially metalated PIX proteins that catalyse reactions that are not catalysed by native Fe-PIX proteins are unknown, and the current, inefficient methods to prepare PIX proteins containing non-native metals have hindered the potential for directed evolution of the resulting enzymes<sup>22–25</sup>.

To evaluate rapidly the potential of artificial [M]-PIX enzymes, we envisioned creating an array of catalysts formed by pairing numerous mutants of apo-PIX proteins and [M]-cofactors in a combinatorial fashion. Previously, apo-PIX proteins have been prepared from native Fe-PIX enzymes by acidic, denaturing extraction of the Fe-cofactor, followed by extensive dialysis to refold the protein<sup>22</sup>. This multistep process is too lengthy for directed evolution, and the harsh, acidic conditions are known to result in proteins that are heterogeneous in structure, which would be detrimental for selective catalysis<sup>26</sup>. Alternatively, Ru-, Mn- and Co-PIX proteins have been expressed directly<sup>23–25</sup>, but these methods are not general, require a gross excess of metal cofactor, and would require a time-consuming purification of each combination of metal and protein. To avoid the aforementioned liabilities of these reported methods in the creation of the proposed catalyst library, we sought to express directly and purify apo-PIX proteins lacking the entire haem unit and to reconstitute them with metal cofactors containing metals other than iron in a stoichiometric fashion (Fig. 1a).

Evaluation of a series of expression conditions revealed those suitable for recombinant expression of the apo-form of haem proteins in *Escherichia coli* (Supplementary Tables 1 and 2). Under the optimized conditions (Supplementary Fig. 1 and Supplementary Table 1), involving minimal media lacking Fe to minimize the biosynthesis of hemin and low temperature to mitigate the instability of the apo-form, we expressed successfully the protein containing less than 5% of the Fe-PIX cofactor, as determined by inductively coupled plasma optical emission spectroscopy (ICP-OES). In particular, mutants of *Physeter macrocephalus* myoglobin (Myo) and *Bacillus megaterium* cytochrome P450 BM3h (P450) with and without an mOCR stability tag were overexpressed in high yields and purified (up to 70 mg l<sup>−1</sup> of protein; Fig. 1a, Supplementary Tables 1 and 2)<sup>9,10,27</sup>. Circular dichroism spectroscopy revealed that these apo-proteins retain the fold of their native Fe-PIX analogues (Fig. 1b, Supplementary Fig. 2). The obtained apo-proteins were reconstituted quantitatively upon addition of stoichiometric amounts

<sup>1</sup>Department of Chemistry, University of California, Berkeley, California 94720, USA. <sup>2</sup>Chemical Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. <sup>3</sup>Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, USA. <sup>4</sup>Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA.

\*These authors contributed equally to this work.



**Figure 1 | Strategy for expedient preparation of [M]-PIX-proteins.** **a**, Direct expression, Ni-NTA purification, and diverse metallation of apo-PIX proteins to generate PIX proteins containing Co, Cu, Mn, Rh, Ir, Ru, and Ag sites or to regenerate the native Fe-containing protein. **b**, Comparison of the circular dichroism spectra obtained from directly expressed apo-Myo, the same protein reconstituted with Fe-PIX (hemin), and the same mutant expressed as a native Fe-PIX protein.

of various [M]-PIX cofactors, as determined by native nano electrospray ionization mass spectrometry (Supplementary Fig. 3). Moreover, reactions catalysed by reconstituted Fe-myoglobin and Fe-P450 occurred with the same enantioselectivities as those catalysed by native Fe-proteins (Supplementary Fig. 4)<sup>9,10</sup>, providing strong evidence that this method indeed generates [M]-PIX-proteins with the intact active site and with the cofactors bound at the native PIX-binding site. Further studies revealed that reconstituted mOCR-myoglobins are stable on storage; reactions catalysed by freshly prepared, frozen, and lyophilized enzymes proceeded with comparable enantioselectivity (see below, Supplementary Fig. 5).

Following this method, we directly expressed eight variants of apo-mOCR-Myo-H93X, each carrying a different mutation to the axial ligand position (H93X). Upon reconstitution of each variant with nine different porphyrin cofactors (containing Fe(Cl)-, Co(Cl)-, Cu-, Mn(Cl)-, Rh-, Ir(Cl)-, Ir(Me)-, Ru(CO)- and Ag-sites, Supplementary Table 2), we rapidly accessed 72 potential catalysts whose activity profiles are distinct from those of wild-type myoglobin, owing to the identity of the metal centre and the amino acid residue serving as the axial ligand (Fig. 2a, b)<sup>3</sup>.

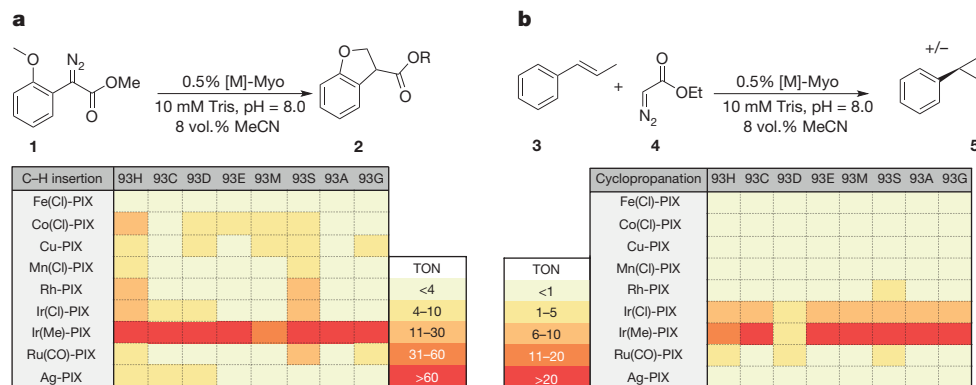
Natural haem proteins functionalize C–H bonds to form C–O bonds<sup>11</sup>, but no haem protein is known to functionalize a C–H bond to form a C–C bond. To identify an enzyme for the insertion of a carbene into a C–H bond, the array of artificial [M]-mOCR-myoglobins containing various metals and axial ligands was evaluated for the reaction of diazoester **1** to form chiral dihydrobenzofuran **2** (Fig. 2a). All myoglobins formed from the native Fe-PIX cofactor were inactive, regardless of the axial ligand. In contrast, non-native metals formed active catalysts when paired with an appropriate axial ligand. The most active catalysts, those containing Ir(Me)-PIX, were formed by incorporating both an abiological metal (Ir) and an abiological axial ligand (–CH<sub>3</sub>) that cannot be incorporated through standard mutagenesis techniques. The eight myoglobins containing Ir(Me)-PIX formed enantioenriched dibenzohydrofuran **2** in up to 50% yield before any further mutagenesis (see below). Moreover, this artificial enzyme tolerated modifications to all portions of the substrate; diazoesters **6–11**, containing varied ester, arene, and alkoxy functionalities (Fig. 3a, Supplementary Fig. 6), also underwent C–H insertion in the presence of Ir(Me)-PIX-Myo. Together, these results show that the

multi-dimensional evaluation of reconstituted PIX-enzymes can identify new artificial metalloenzymes that catalyse reactions that biological Fe-PIX-proteins do not catalyse.

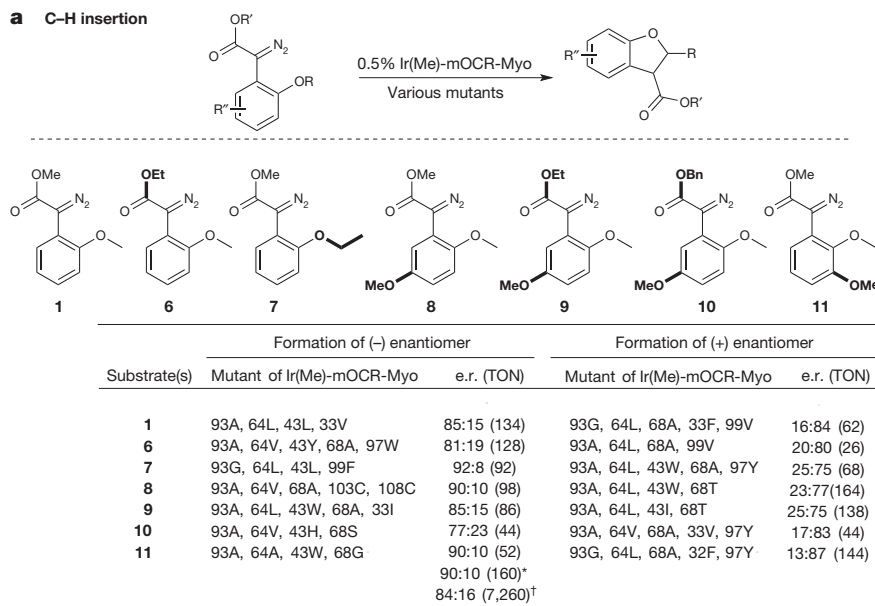
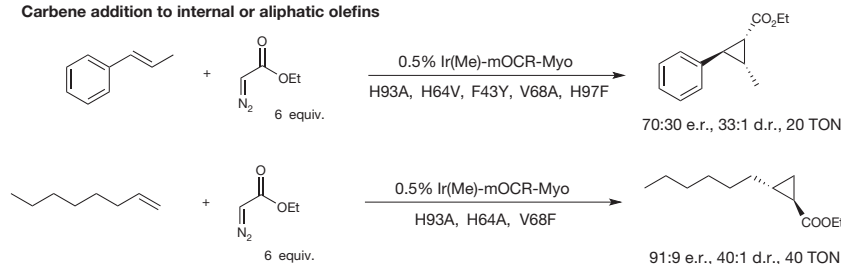
A substantial benefit of using enzymes for synthetic applications is the potential to use directed evolution to obtain a catalyst with desired properties<sup>13,28</sup>. To develop an enantioselective catalyst for each of the seven substrates undergoing C–H insertion, we followed a hybrid strategy based on stepwise optimization of small sets of amino acids progressively more distal from the reaction site (Fig. 4). In the first phase, the axial ligand (H93) was modified to A or G and the residue directly above the metal centre (H64) was modified to A, V, L, or I (Fig. 4) to give an initial set of eight mutants. In the second phase, these initial eight mutants were modified at positions F43 and V68, which are located in the binding site (Fig. 4), to generate 225 prospective enzymes. To retain the hydrophobicity of the site that binds the porphyrin and substrate, only hydrophobic and uncharged residues (V, A, G, F, Y, S, T) were introduced at positions F43 and V68. Of these 225 mutants, 22 that were among the most selective for one or more of the substrates in Fig. 3 were subjected to a further round of evolution during which the residues at four additional positions (L32, F33, H97, and I99) were modified to generate 217 more mutants (Fig. 4).

The complete results of the carbene insertion reaction with these mutants are provided in Supplementary Tables 5–11 and are summarized in Fig. 3b and Supplementary Table 4. The directed evolution of Ir(Me)-myoglobins uncovered distinct enzymes catalysing the C–H functionalization to form either enantiomer of the products containing a new C–C bond formed from substrates **1** and **6–11**. The reactions occurred with selectivities up to an enantiomeric ratio (e.r.) of 92:8 and with yields up to 97% (Fig. 4, Supplementary Fig. 7 and Supplementary Tables 4–11) with enzymes that were evolved from those giving nearly racemic product. The Ir-myoglobins are suitable catalysts for synthetic-scale reactions; the carbene insertion of substrate **11** formed the product containing a new C–C bond in 80% isolated yield from a reaction of 28 mg of **11** with nearly the same enantioselectivity as observed on smaller scale (Fig. 3). A reaction conducted with a 40,000:1 ratio of substrate to Ir(Me)-mOCR-myoglobin occurred with a turnover number (TON) of 7,200 (Fig. 3).

In contrast to the few directed evolutions of artificial enzymes reported previously<sup>28</sup>, our method of preparing variants of the

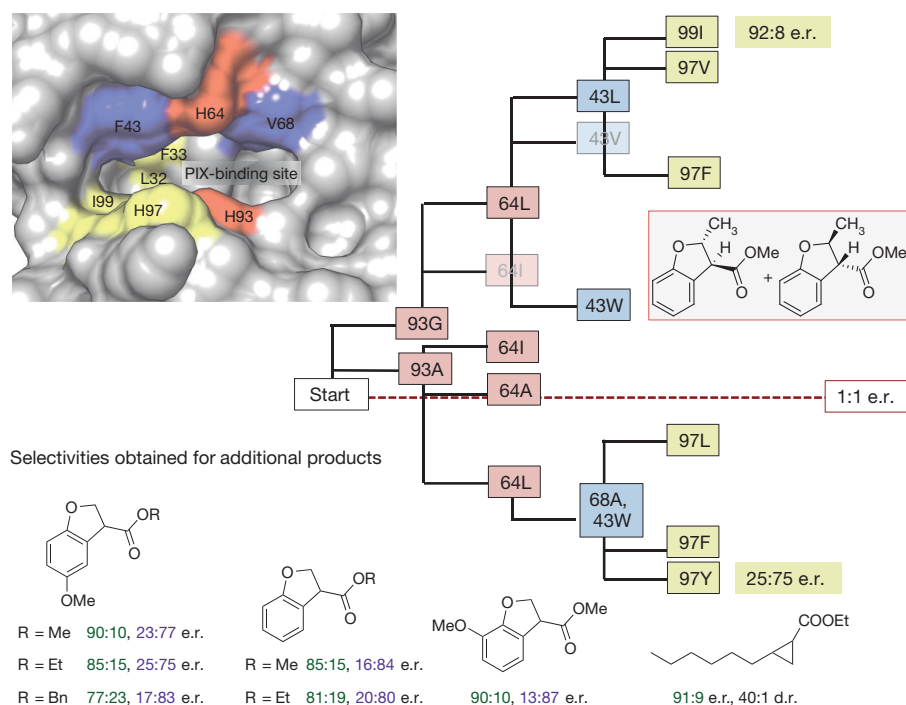


**Figure 2 | Evaluation of artificial [M]-PIX mOCR-myoglobins as catalysts.** **a**, **b**, Catalysts for the insertion of carbenes into C–H bonds (**a**; **1** → **2**) and for the addition of carbenes to an internal olefin (**b**; **3** + **4** → **5**). TON, turnover number. Reaction conditions for the C–H insertion reaction: 10 mM substrate and 0.5% catalyst in 250 μl buffer (10 mM Tris, pH 8.0 containing 8 vol.% MeCN). Reaction conditions for the cyclopropanation reaction: 10 mM olefin, 30 mM EDA, and 0.5% catalyst in 250 μl buffer (10 mM Tris, pH 8.0 containing 8 vol.% MeCN).

**a C–H insertion****b Carbene addition to internal or aliphatic olefins**

Ir(Me)-PIX-enzyme enabled us to pursue an individual, eight-site evolutionary trajectory for the reaction of each substrate that identified catalysts selectively forming either enantiomer of all targeted products. These results demonstrate that Ir(Me)-PIX-myoglobins are highly evolvable for different substrates containing varied structural modifications. These results, along with the high isolated yield and

the observation of high TONs, demonstrate that the direct expression of apo-Myo, the insertion of diverse [M]-PIX cofactors, and the subsequent directed evolution of the most active enzymes identified is a robust strategy that can be applied in a general way to create stereo-selective enzymes for abiological catalysis that cannot be accomplished by any natural enzymes.



**Figure 3 | Summary of activities and selectivities for reactions catalysed by Ir(Me)-mOCR-Myo.**

**a**, Substrates (1, 6–11) for C–H insertion reactions and the most selective mutants identified through the directed evolution. Notations (+)– and (–)–enantiomer distinguish the formation of opposite enantiomers of the product. \*Reaction on a 0.12 mmol scale of substrate, TON based on isolated yield of product (80%). †Reaction on a 0.05 mmol scale of substrate with 0.0025% catalyst. Reaction conditions: 10 mM substrate and 0.5% catalyst in 250 µl buffer (10 mM Tris, pH 8.0 containing 8 vol.% MeCN). **b**, Carbene addition to internal and aliphatic olefins catalysed by mutants of Ir(Me)-mOCR Myo that were found to be the most selective. d.r., diastereomeric ratio. Reaction conditions: 10 mM olefin, 60 mM EDA, and 0.5% catalyst in 250 µl buffer (10 mM Tris, pH 8.0 containing 8 vol.% MeCN). EDA was added over 12 h via syringe pump.

**Figure 4 | Directed evolution strategy used to obtain Ir(Me)-mOCR-Myo mutants**

capable of producing either enantiomer of the products of C–H insertion reactions of varied substrates. Inner sphere (red boxes), middle sphere (blue), and outer sphere (yellow) residues are highlighted in the depiction of the active site (top left) and in the evolutionary tree. (Image of active site and its surroundings produced in Chimera from PDB 1MBN<sup>30</sup>.) The strategy is exemplified by showing the enantioselectivities achieved in the formation of the product boxed on the right. Mutants positioned above the red dotted line formed predominantly the opposite enantiomer of those shown below the dotted red line. Some mutants are shown in boxes that are semi-transparent for clarity of the figure. In the case of selectivities obtained for additional products, the e.r. values given in green and purple colour are those for reactions forming predominantly the opposite enantiomers. The reactions were run under the conditions described in Fig. 3.

To assess the generality of this approach further, we sought catalysts for the cyclopropanation of internal alkenes and  $\alpha$ -olefins that, like carbene insertion into C–H bonds, have not been accomplished with natural or artificial enzymes. As a starting point, we evaluated the two-dimensional array of [M]-PIX catalysts shown in Fig. 2b for the cyclopropanation of  $\beta$ -methylstyrene **3** with ethyl diazoacetate **4** (EDA). In agreement with literature reports<sup>10</sup>, Fe-PIX enzymes did not catalyse this reaction (Fig. 2b). In contrast, Rh-, Ru-, and Ir-PIX enzymes furnished the cyclopropane product **5**. The enzyme containing Ir(Me)-PIX was the most active. Although further work is needed to obtain full conversion and high enantiomeric excess (e.e.), the reaction of EDA with  $\beta$ -methylstyrene catalysed by the Ir(Me)-mOCR-Myo mutant H93A, H64V, F43Y, V68A, H97F formed the cyclopropane **5** with a TON of 40 with 70:30 e.r., and with a high >33:1 ratio of diastereomers, favouring the *trans* isomer.

Having observed the expanded scope of enzyme-catalysed cyclopropanation, we assessed the ability of Ir(Me)-PIX enzymes to catalyse the cyclopropanation of 1-octene, an unactivated, aliphatic olefin. The series of Ir(Me)-PIX-Myo enzymes assessed for C–H insertion reactions were tested as catalysts for the reaction of EDA with 1-octene. Although the mutant H93A, H64A, V68F formed the products of C–H insertion from all substrates unselectively, the same mutant formed the product of cyclopropanation of 1-octene in an enantiomeric ratio of 91:9 and a *trans:cis* ratio of 40:1 (Figs 3b, 4; Supplementary Tables 12–14). Cyclopropanations of aliphatic alkenes catalysed by traditional metal complexes are typically conducted with an excess of the alkene<sup>29</sup>. In contrast, the Ir(Me)-PIX-Myo mutant catalyses the reaction with an excess of EDA (a TON of 42 with a 10:1 ratio of EDA:1-octene), suggesting that reactions can be developed with valuable alkenes as limiting reagent. The reactions with fewer equivalents of EDA occur with lower TON due to consumption of EDA by dimerization or O–H insertion of water. These cyclopropanations of unactivated alkenes show the broad potential to evolve artificial myoglobins containing abiological active sites for reactions that are not catalysed by enzymes containing native metals.

The work presented here demonstrates that unknown enzymatic reactivity can be achieved by incorporating just a metal ion with an accompanying small ligand into a well-known metalloprotein, while retaining the native structure of the active site. Selectivity for specific substrates, then, can be achieved readily by directed evolution. Considering the rich chemistry of free metalloporphyrins and the ease of preparation and evolution of haem proteins containing diverse metals by the methods just described, this methodology should seed the creation of many new artificial metalloenzymes with diverse, unnatural reactivity. Moreover, the facile, direct expression of apo-haem proteins could be used in tandem with strategies to incorporate highly active noble-metal complexes of ligands beyond porphyrins. Access to such a range of artificial haem proteins provides a nearly limitless opportunity to achieve catalytic reactions with selectivity derived from the interaction of the substrate with a natural, evolvable binding site.

Received 18 November 2015; accepted 15 March 2016.

Published online 13 June 2016.

- Lewis, J. C. Artificial metalloenzymes and metalloprotein catalysts for organic synthesis. *ACS Catal.* **3**, 2954–2975 (2013).
- Farwell, C. C., Zhang, R. K., McIntosh, J. A., Hyster, T. K. & Arnold, F. H. Enantioselective enzyme-catalyzed aziridination enabled by active-site evolution of a cytochrome P450. *ACS Cent. Sci.* **1**, 89–93 (2015).
- Hyster, T. K. & Arnold, F. H. P450<sub>BM3</sub>-axial mutations: a gateway to non-natural reactivity. *Isr. J. Chem.* **55**, 14–20 (2015).
- Ringenberg, M. R. & Ward, T. R. Merging the best of two worlds: artificial metalloenzymes for enantioselective catalysis. *Chem. Commun.* **47**, 8470–8476 (2011).
- Ward, T. R. Artificial metalloenzymes based on the biotin–avidin technology: enantioselective catalysis and beyond. *Acc. Chem. Res.* **44**, 47–57 (2011).
- Jing, Q. & Kazlauskas, R. J. Regioselective hydroformylation of styrene using rhodium-substituted carbonic anhydrase. *ChemCatChem* **2**, 953–957 (2010).
- Abe, S. *et al.* Polymerization of phenylacetylene by rhodium complexes within a discrete space of apo-ferritin. *J. Am. Chem. Soc.* **131**, 6958–6960 (2009).

- Key, H. M., Clark, D. S. & Hartwig, J. F. Generation, characterization, and tunable reactivity of organometallic fragments bound to a protein ligand. *J. Am. Chem. Soc.* **137**, 8261–8268 (2015).
- Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307–310 (2013).
- Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744–1748 (2015).
- Ortiz de Montellano, P. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev.* **110**, 932–948 (2010).
- Peters, M. W., Meinhold, P., Glieder, A. & Arnold, F. H. Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *J. Am. Chem. Soc.* **125**, 13442–13450 (2003).
- Roiban, G.-D. & Reetz, M. T. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.* **51**, 2208–2224 (2015).
- Tyagi, V., Bonn, R. B. & Fasan, R. Intermolecular carbene S–H insertion catalyzed by engineered myoglobin-based catalysts. *Chem. Sci.* **6**, 2488–2494 (2015).
- Chan, K. H., Guan, X., Lo, V. K. Y. & Che, C.-M. Elevated catalytic activity of ruthenium(II)–porphyrin-catalyzed carbene/nitrene transfer and insertion reactions with N-heterocyclic carbene ligands. *Angew. Chem. Int. Ed.* **53**, 2982–2987 (2014).
- Maxwell, J. L., O'Malley, S., Brown, K. C. & Kodadek, T. Shape-selective and asymmetric cyclopropanation of alkenes catalyzed by rhodium porphyrins. *Organometallics* **11**, 645–652 (1992).
- Anding, B. J., Ellern, A. & Woo, L. K. Olefin cyclopropanation catalyzed by iridium(III) porphyrin complexes. *Organometallics* **31**, 3628–3635 (2012).
- Carey, J. R. *et al.* A site-selective dual anchoring strategy for artificial metalloprotein design. *J. Am. Chem. Soc.* **126**, 10812–10813 (2004).
- Ohashi, M. *et al.* Preparation of artificial metalloenzymes by insertion of chromium(III) Schiff base complexes into apomyoglobin mutants. *Angew. Chem. Int. Ed.* **42**, 1005–1008 (2003).
- Oohora, K., Kihira, Y., Mizohata, E., Inoue, T. & Hayashi, T. C(sp<sup>3</sup>)–H bond hydroxylation catalyzed by myoglobin reconstituted with manganese porphyrine. *J. Am. Chem. Soc.* **135**, 17282–17285 (2013).
- Bordeaux, M., Singh, R. & Fasan, R. Intramolecular C(sp<sup>3</sup>)–H amination of arylsulfonfyl azides with engineered and artificial myoglobin-based catalysts. *Bioorg. Med. Chem.* **22**, 5697–5704 (2014).
- Teale, F. W. Cleavage of the haem-protein link by acid methylethylketone. *Biochim. Biophys. Acta* **35**, 543 (1959).
- Lelyveld, V. S., Brustad, E., Arnold, F. H. & Jasanoff, A. Metal-substituted protein MRI contrast agents engineered for enhanced relaxivity and ligand sensitivity. *J. Am. Chem. Soc.* **133**, 649–651 (2011).
- Kawakami, N., Shoji, O. & Watanabe, Y. Single-step reconstitution of apohemoproteins at the disruption stage of *Escherichia coli* cells. *ChemBioChem* **13**, 2045–2047 (2012).
- Woodward, J. J., Martin, N. I. & Marletta, M. A. An *Escherichia coli* expression-based method for heme substitution. *Nat. Methods* **4**, 43–45 (2007).
- Paulson, D. R., Addison, A. W., Dolphin, D. & James, B. R. Preparation of ruthenium(II) and ruthenium(III) myoglobin and the reaction of dioxygen, and carbon monoxide, with ruthenium(II) myoglobin. *J. Biol. Chem.* **254**, 7002–7006 (1979).
- DelProposto, J., Majmudar, C. Y., Smith, J. L. & Brown, W. C. Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr. Purif.* **63**, 40–49 (2009).
- Ilie, A. & Reetz, M. T. Directed evolution of artificial metalloenzymes. *Isr. J. Chem.* **55**, 51–60 (2015).
- Suematsu, H., Kanchiku, S., Uchida, T. & Katsuki, T. Construction of aryliridium-salen complexes: enantio- and cis-selective cyclopropanation of conjugated and nonconjugated olefins. *J. Am. Chem. Soc.* **130**, 10327–10337 (2008).
- Watson, H. C. The stereochemistry of the protein myoglobin. *Protein Stereochem.* **4**, 299 (1969).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** This work was supported by the Director, Office of Science, of the US Department of Energy under contract no. DE-AC02-05CH11231, by the NSF (graduate research fellowship to H.M.K.), and the NWO Netherlands Organization for Scientific Research (Rubicon postdoctoral fellowship no. 680-50-1306 to P.D.). We thank the QB3 MacroLab facility (sub-cloning), the UC Berkeley DNA Sequencing Facility (plasmid sequencing), T. Iavarone and the QB3 Mass Spectrometry Facility (supported by NIH grant 1S10R022393-01) for native NS-ESI-MS data and analysis, and H. Zhao (University of Illinois-Champaign Urbana) for the P411-CIS gene.

**Author Contributions** H.M.K., P.D., and J.F.H. conceived the work, designed the initial experiments, and interpreted the data. All authors discussed the results to design subsequent experiments. H.M.K. and P.D. performed all the experiments. H.M.K., P.D., and J.F.H. wrote the manuscript and all authors reviewed the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F.H. ([jhartwig@berkeley.edu](mailto:jhartwig@berkeley.edu)).

# A complement–microglial axis drives synapse loss during virus–induced memory impairment

Michael J. Vasek<sup>1</sup>, Charise Garber<sup>1</sup>, Denise Dorsey<sup>1</sup>, Douglas M. Durrant<sup>1,2</sup>, Bryan Bollman<sup>1</sup>, Allison Soung<sup>1</sup>, Jinsheng Yu<sup>3</sup>, Carlos Perez-Torres<sup>4</sup>, Arnaud Frouin<sup>5</sup>, Daniel K. Wilton<sup>5</sup>, Kristen Funk<sup>1</sup>, Bette K. DeMasters<sup>6</sup>, Xiaoping Jiang<sup>7</sup>, James R. Bowen<sup>8</sup>, Steven Mennerick<sup>7</sup>, John K. Robinson<sup>9</sup>, Joel R. Garbow<sup>4</sup>, Kenneth L. Tyler<sup>6</sup>, Mehul S. Suthar<sup>8</sup>, Robert E. Schmidt<sup>10</sup>, Beth Stevens<sup>5</sup> & Robyn S. Klein<sup>1,10,11</sup>

**Over 50% of patients who survive neuroinvasive infection with West Nile virus (WNV) exhibit chronic cognitive sequelae<sup>1,2</sup>. Although thousands of cases of WNV-mediated memory dysfunction accrue annually<sup>3</sup>, the mechanisms responsible for these impairments are unknown. The classical complement cascade, a key component of innate immune pathogen defence, mediates synaptic pruning by microglia during early postnatal development<sup>4,5</sup>. Here we show that viral infection of adult hippocampal neurons induces complement-mediated elimination of presynaptic terminals in a murine WNV neuroinvasive disease model. Inoculation of WNV-NS5-E218A, a WNV with a mutant NS5(E218A) protein<sup>6,7</sup> leads to survival rates and cognitive dysfunction that mirror human WNV neuroinvasive disease. WNV-NS5-E218A-recovered mice (recovery defined as survival after acute infection) display impaired spatial learning and persistence of phagocytic microglia without loss of hippocampal neurons or volume. Hippocampi from WNV-NS5-E218A-recovered mice with poor spatial learning show increased expression of genes that drive synaptic remodelling by microglia via complement. C1QA was upregulated and localized to microglia, infected neurons and presynaptic terminals during WNV neuroinvasive disease. Murine and human WNV neuroinvasive disease post-mortem samples exhibit loss of hippocampal CA3 presynaptic terminals, and murine studies revealed microglial engulfment of presynaptic terminals during acute infection and after recovery. Mice with fewer microglia (*Il34*<sup>−/−</sup> mice with a deficiency in IL-34 production) or deficiency in complement C3 or C3a receptor were protected from WNV-induced synaptic terminal loss. Our study provides a new murine model of WNV-induced spatial memory impairment, and identifies a potential mechanism underlying neurocognitive impairment in patients recovering from WNV neuroinvasive disease.**

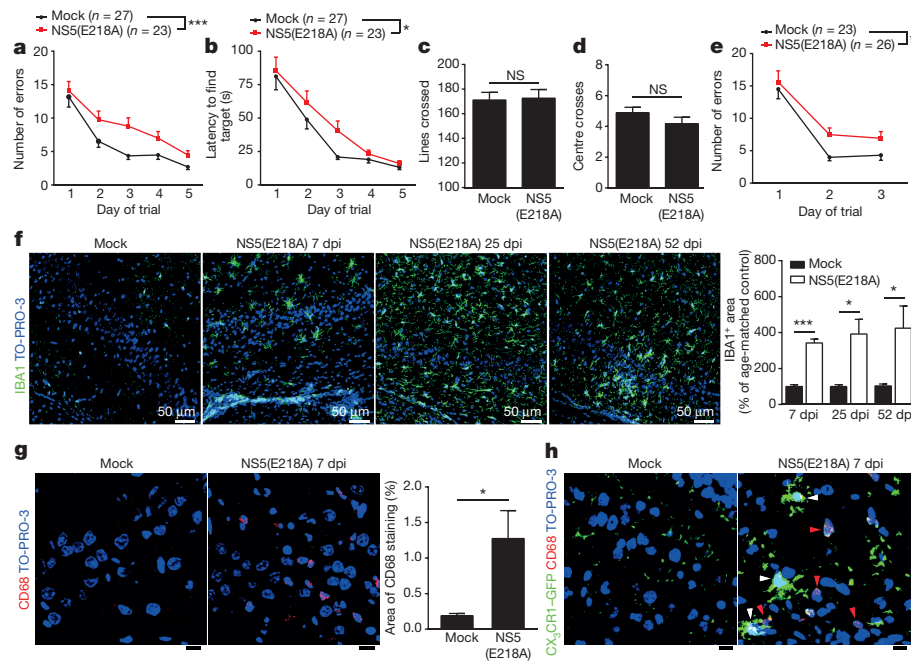
Studies in humans and rodents indicate that WNV targets neurons within the hippocampus<sup>8</sup>, which is essential for spatial and contextual memory formation<sup>9</sup>. Patients that survive WNV neuroinvasive disease often exhibit impaired visuospatial processing and memory<sup>1,10</sup>. In post-mortem samples of patients with WNV neuroinvasive disease and in rodent models with low survival rates (<50%), significant neuronal loss, inflammation, and microglial activation occur within infected brain regions<sup>11–14</sup>. However, the extent of viral burden and neuron loss may be much lower in individuals who survive WNV neuroinvasive disease and may vary between brain regions. In addition, host–pathogen interactions could explain the range of cognitive sequelae experienced by WNV neuroinvasive disease survivors<sup>1</sup>.

Mechanisms underlying cognitive impairments in patients recovering from West Nile neuroinvasive disease are unknown, mainly due to lack of murine recovery models. Current models that use virulent WNV strains yield either 100% death following intracranial WNV infection or 10–70% survival and variable CNS viral burdens following peripheral routes of infection<sup>15</sup>. To circumvent this in order to develop a model of recovery from WNV neuroinvasive disease, we used a strain of WNV with a point mutation in nonstructural protein 5 NS5(E218A), which lacks functional 2′-O methyltransferase that generates a 5′ cap on viral RNA to evade type I interferon-mediated restriction of viral translation<sup>6,16</sup>. While WNV-NS5-E218A replicates in permissive cells, 90% of 8-week-old mice survive intracranial inoculation (Extended Data Fig. 1a), with uniform viral brain burdens peaking between 6–8-days post-infection (dpi)<sup>7</sup>, followed by viral clearance at 15 dpi (Extended Data Fig. 1b, c). Intracranial infection with WNV-NS5-E218A induces neuroinflammation within 7 dpi, with numbers and phenotypes of infiltrating leukocytes similar to intracranial infection with virulent strain WNV-NY99 (Extended Data Fig. 1d), and consistent with data demonstrating early reversion of WNV-NS5-E218A to wild-type virus within the central nervous system<sup>7</sup>. WNV-NS5-E218A intracranial infection results in few apoptotic neurons (0.5–1.5%) at 7 dpi (Extended Data Fig. 1e), similar to 8-week-old mice after footpad infection with WNV-NY99 (ref. 12). Thus, WNV-NS5-E218A may be used to examine behavioural, cellular, and molecular mechanisms of recovery from WNV neuroinvasive disease.

The Barnes maze behavioural task was used to determine whether WNV-NS5-E218A-recovered mice exhibit neurocognitive deficits<sup>17</sup>. At 46 dpi, WNV-infected mice exhibit slower learning, commit more errors (Fig. 1a) and require more time (Fig. 1b) before locating the target hole than mock-infected controls (Supplementary Videos 1 and 2). Studies performed in mice at 22 dpi showed similar effects (Fig. 1e and Extended Data Fig. 1f). WNV-NS5-E218A-recovered animals improve over the 5 days of training, although they continue to make significantly more errors than controls on each day. Recovered animals do not exhibit impairments in motor activity or exploratory anxiety in an open-field behavioural assessment at 45 dpi (Fig. 1c, d) or 21 dpi (Extended Data Fig. 1g), indicating that impairments in Barnes maze solving are specific to spatial learning.

Caspase-3-dependent apoptosis of neurons occurs in animals that succumb to WNV neuroinvasive disease<sup>12</sup>. TUNEL and NeuN-staining did not detect ongoing apoptosis (Extended Data Fig. 2a) or loss of NeuN<sup>+</sup> neurons within the circuitry of the entorhinal cortex and hippocampus that serve spatial learning (Extended Data Fig. 2b).

<sup>1</sup>Department of Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>2</sup>Biological Sciences Department, California State Polytechnic University, 3801 West Temple Avenue, Pomona, California 91768, USA. <sup>3</sup>Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>4</sup>Department of Radiology, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>5</sup>Department of Neurology, F. M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>6</sup>Department of Neurology, University of Colorado School of Medicine, Aurora, Colorado 80045, USA. <sup>7</sup>Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>8</sup>Department of Pediatrics and Children's Healthcare of Atlanta, Emory Vaccine Center, Yerkes National Primate Research Center, Emory University School of Medicine, Atlanta, Georgia 30329, USA. <sup>9</sup>Department of Psychology, Stony Brook University, Stony Brook, New York 11794, USA. <sup>10</sup>Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>11</sup>Department of Anatomy and Neurobiology, Washington University School of Medicine, St Louis, Missouri 63110, USA.



**Figure 1 | WNV-mediated spatial learning and memory impairments and activated microglia persist beyond 45 days post-infection.** **a, b,** At 46 days post-infection (dpi), mock or WNV-NS5-E218A-infected mice underwent 5 days of testing on the Barnes maze spatial learning task. Errors (**a**) and latency (**b**) before finding target hole were scored daily (mean of 2 trials per day, \*\*\* $P < 0.001$ , \* $P < 0.05$  by repeated measures two-way ANOVA). **c, d,** At 45 dpi, mice were observed on the open-field test and assessed for locomotor activity (**c**) and anxiety (**d**). **a–d,** Mock ( $n = 27$ ) and WNV-NS5-E218A-infected ( $n = 23$ ) mice. **e,** Mock ( $n = 23$ ) and WNV-NS5-E218A-infected ( $n = 26$ ) mice were tested at 22 dpi on a 3-day version of the Barnes maze, and evaluated as in **a**. **f,** Immunostaining for IBA1 in control and WNV-NS5-E218A-infected mice at 7 dpi ( $n = 6$  or 7 per group for control or WNV, respectively), 25 dpi ( $n = 3$  or 4 for

control or WNV, respectively), and 52 dpi ( $n = 6$  or 4 for control or WNV, respectively) (mean of 2 technical replicates used). **g, h,** Immunostaining shows increased levels of CD68, a microglial/macrophage lysosomal activation marker, in WNV-NS5-E218A-infected wild-type mice (**g**) ( $n = 4$  mice per group) and CX<sub>3</sub>CR1-GFP<sup>+/−</sup> (**h**) ( $n = 3$  mice per group) mice. **h,** CD68 is present within CX<sub>3</sub>CR1-positive microglia (white arrowheads) and infiltrating macrophages (red arrowheads). Images are representative of at least 3 mice per group. All panels, \*\*\* $P < 0.001$ , \* $P < 0.05$ , NS, not significant by two-tailed  $t$ -test, and scale bars, 10 μm, unless otherwise noted. Error bars, s.e.m. Immunostaining and quantification were performed within the hippocampal CA3 region.

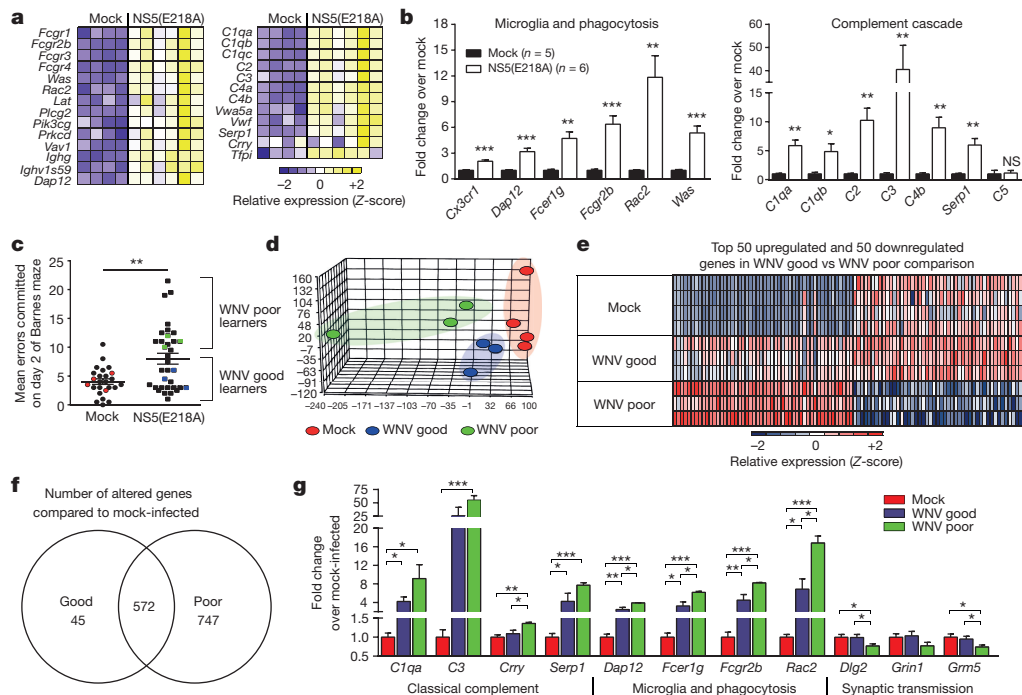
Similarly, hippocampal and total brain volumes at 52 dpi do not differ between mock-infected and WNV-NS5-E218A-infected animals (Extended Data Fig. 2c), and GFAP staining within the hippocampus was unchanged (Extended Data Fig. 2d). However, microglial nodules were detected within the hippocampus of WNV-NS5-E218A-recovered animals at 7, 25, and 52 dpi (Extended Data Fig. 3b). These nodules contained increased numbers of IBA1-positive cells with activated cell morphology (Fig. 1e). At 7 dpi, increased levels of CD68, a marker of microglial/macrophage lysosomal activation, were observed within both microglia and infiltrating macrophages of CX<sub>3</sub>CR1-GFP<sup>+/−</sup> mice (Fig. 1g, h), however macrophage infiltration was absent by 25 dpi (Extended Data Fig. 2f).

Whole-transcriptome microarray of hippocampi RNA from mock-infected versus WNV-NS5-E218A-infected mice at 25 dpi enabled detection of differential expression of 1,364 transcripts. Pathway analysis identified signatures associated with the generation and maintenance of synapses, activation of innate immune responses, and microglial proteins involved in sensing endogenous ligands and microbes<sup>18</sup> (Supplementary Table 1). In the latter category, we identified genes associated with microglial-mediated phagocytosis (*Cx3cr1* which encodes CX<sub>3</sub>CR1; *Dap12* which encodes Dap12 (also known as Tyrobp); *Fcgr1g* which encodes FcγR1G; *Fcgr2b* which encodes FcγR2; *Rac2* which encodes Rac2 and *Was* which encodes WAS) and the classical complement pathway (*C1qa* which encodes C1QA; *C2* which encodes C2; *C3* which encodes C3; *C4b* which encodes C4b and *Serp1* which encodes Serpin1) (Fig. 2a), which were validated using quantitative PCR (qPCR) (Fig. 2b). C1q and C3 are required for retinogeniculate and cortical synaptic pruning during murine CNS development<sup>4,5,19</sup>. Although complement contributes to control

of WNV dissemination following peripheral infection<sup>20</sup>, complement expression within the brain during WNV neuroinvasive disease has not been investigated.

To identify genetic signatures specific to spatial learning defects, we categorized WNV-NS5-E218A-recovered mice into those that perform similar to mock-infected animals (good learners categorized as <8 errors) and those that exhibit severe learning deficits (poor learners categorized as >9.5 errors), on day 2 of Barnes maze testing (Figs 1e and 2c–e). A total of 747 genes were increased in the WNV poor learners, 45 genes altered in WNV good learners, and 572 genes altered in both groups compared to mock-infected littermates (Fig. 2f) (Supplementary Tables 2 and 3). WNV-NS5-E218A-recovered animals with poor learning exhibited increased levels of CRRY (also known as Cr11), Dap12, FcγR2, Rac2, and C1QA and decreased levels of Dlg2, a synaptic scaffolding protein, and the metabotropic glutamate receptor, Grm5, compared with good learners and mock-infected animals (Fig. 2g). Grm5 is downregulated in mouse brain during acute WNV, Japanese encephalitis virus, and reovirus infections<sup>21</sup>. KEGG pathway analysis of upregulated genes in WNV poor memory compared to WNV good memory mice revealed top pathways of cytokine signalling, calcium signalling, and B-cell receptor signalling (Supplementary Table 4), whereas top pathways of downregulated genes include long-term potentiation, axon guidance, and Wnt signalling (Supplementary Table 5).

Loss in neurons or persistence of virus could contribute to cognitive dysfunction. Evaluation of neuronal numbers throughout the hippocampus and entorhinal cortices of mock-infected and WNV-infected good and poor learners revealed no differences (Extended Data Fig. 2b). Analyses of WNV envelope protein positive and negative



**Figure 2 | Transcriptional profile of good and poor spatial learners during WNV recovery.** **a**, Heat maps show relative expression of significantly altered genes (see Methods) generated from hippocampal microarray of mock vs WNV-NS5-E218A-recovered mice at 25 dpi, each column represents individual mice. **b**, Validation of select genes and pathways in a unique set of mice by qPCR (mock ( $n = 5$ ) and WNV-NS5-E218A ( $n = 6$ ) mice). **c**, Scatter plot depicting number of errors committed on day 2 of Barnes maze testing, showing good (blue) and poor (green) learners among WNV-NS5-E218A-infected and mock-infected (red) controls. **d**, Principle component analysis of microarray samples separated by groups as in **c**. WNV, West Nile virus. **e**, Relative expression

heat map showing the top 50 upregulated and 50 downregulated genes by microarray comparing WNV-recovery good and poor learners; each row represents individual mice. **f**, Venn diagram of microarray data showing number of genes significantly altered from mock-infected controls ( $P < 0.05$ , fold change  $> 1.5$ ) in WNV-recovery good or poor learners. **g**, Validation by qPCR of select genes altered between WNV good learners and WNV poor learners using a separate cohort of mice (mock ( $n = 5$ ), WNV good ( $n = 3$ ), and WNV poor ( $n = 3$ ) mice). All panels, \*\*\* $P < 0.001$ , \* $P < 0.05$ , NS, not significant by two-tailed  $t$ -test. Error bars, s.e.m.

strand RNA from the hippocampi of WNV-NS5-E218A-recovered mice showed high levels of both strands at 7 dpi, which decreased by 25 and 52 dpi, with no differences in levels of either strand between good and poor learners at either time point (Extended Data Fig. 1h–j). These data suggest that persistence of replicating WNV-NS5-E218A does not contribute to alterations in learning.

Given the alterations in genes related to synaptic function, we quantified synaptic terminals within the hippocampus of WNV-NS5-E218A-infected mice. Numbers of colocalized presynaptic and postsynaptic puncta within the stratum lucidum of the hippocampal CA3 (mossy fibre terminals) were decreased at 7 dpi in WNV-NS5-E218A-infected animals compared to mock-infected controls (Fig. 3a). The decrease in colocalization was traced to a 40% reduction in number, but not size, of presynaptic terminals (Extended Data Fig. 3a), with no change in numbers of postsynaptic terminals (Fig. 3a). Altered expression of the presynaptic glutamatergic vesicular transporter, VGLUT1 (also known as Slc17a7), in the hippocampus has been linked to cognitive impairment in rodents<sup>22</sup>. Evidence of hippocampal glutamatergic synapse loss was detectable at 25 dpi and WNV-NS5-E218A-recovered mice with poor spatial learning exhibited fewer VGLUT1-positive synaptic puncta than WNV-NS5-E218A-infected good learners, which were fewer than in mock-infected controls (Fig. 3b). WNV-NY99 infection led to similar reductions in synaptic terminals (Extended Data Fig. 3b). Of note, mock-infected and WNV-NS5-E218A-recovered mice (all learners) display similar levels of phosphorylated neurofilament heavy chain (SMI-31) within hippocampal mossy fibre tracts (Extended Data Fig. 3c). These results indicate that synapse elimination does not lead to neuronal death.

Acute post-mortem WNV neuroinvasive disease patient specimens similarly show reduced numbers of CA3 presynaptic terminals

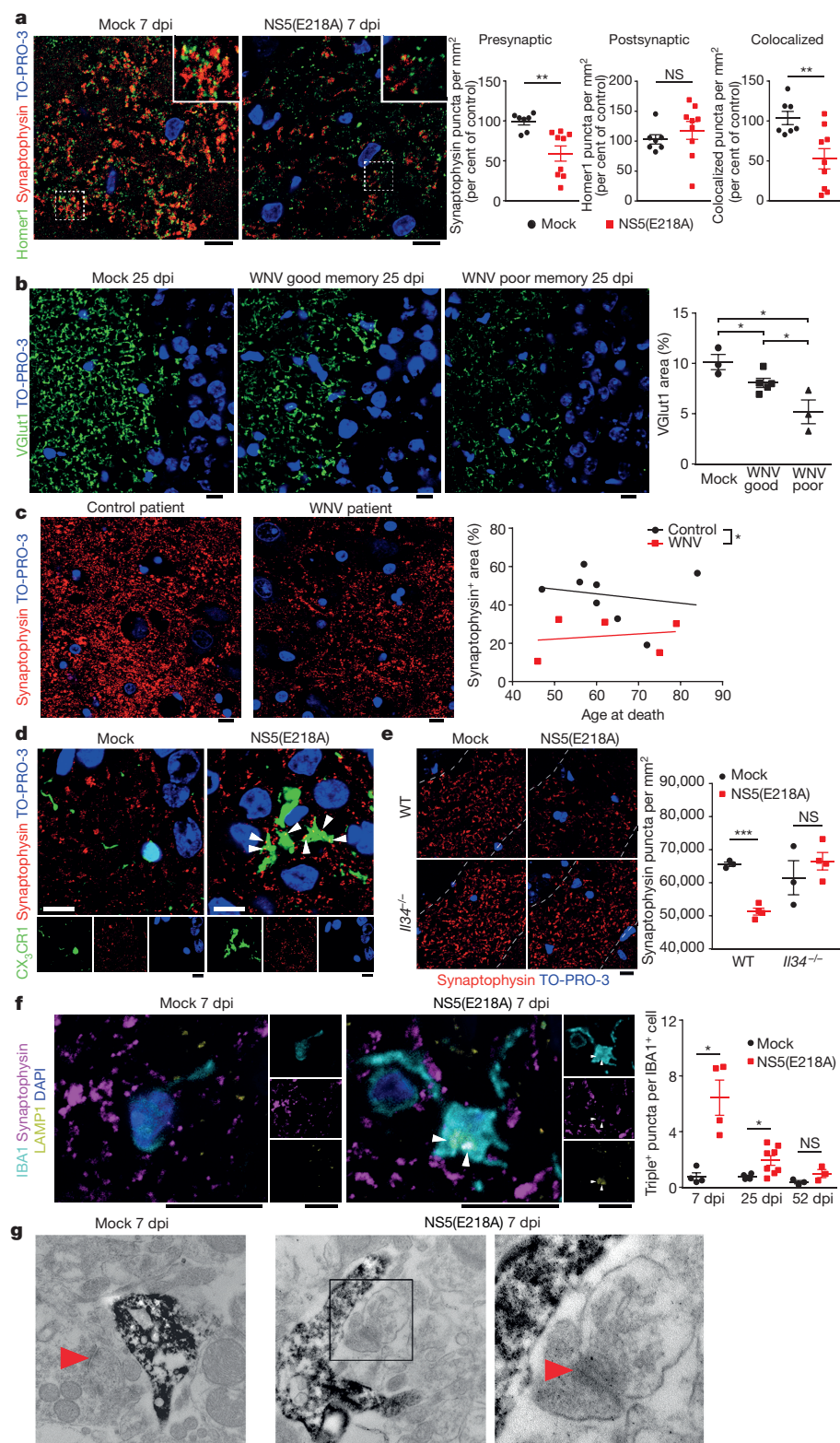
compared to age-matched patient controls (Fig. 3c and Supplementary Table 5). In 3 out of 5 WNV cases, WNV antigen was detected within the CA2/CA3 region, but absent from neighbouring regions (Supplementary Table 6 and Extended Data Fig. 4a). Hippocampal CA1 and entorhinal cortex in the human samples also displayed presynaptic terminal loss (Extended Data Fig. 4b, c). These data show that synaptic pathology is observed in areas without detectable viral antigen. Thus, WNV infection may alter synapse homeostasis both in affected and connected brain regions.

Given their phagocytic appearance in WNV-recovered mice, we wondered if microglia were driving synapse loss. Three-dimensional reconstructions of microglia from CX3CR1-GFP<sup>+/+</sup> WNV-NS5-E218A-infected mice (Fig. 3d) revealed synaptophysin-positive puncta within GFP<sup>+</sup> cells (Supplementary Videos 3 and 4). Additionally, presynaptic terminals from mice with fewer and less proliferative microglia (*IL34*<sup>−/−</sup>)<sup>23</sup> were not eliminated during WNV-NS5-E218A acute infection (Fig. 3e). Colocalization of synaptophysin, lysosomal-associated membrane protein 1 (LAMP1), and IBA1, revealed increased numbers of presynaptic terminals within the lysosomes of IBA1-positive cells, but not S100 $\beta$ <sup>+</sup> astrocytes (Extended Data Fig. 3d), in the CA3 of WNV-NS5-E218A-infected mice at 7 and 25 dpi compared with mock-infected animals (Fig. 3f). Numbers of engulfed presynaptic terminals recover to baseline levels by 52 dpi, suggesting that presynaptic elimination eventually abates. Electron microscopy of the CA3 and molecular layer (Fig. 3g) during acute WNV neuroinvasive disease identified microglia enriched in phagosomes (Extended Data Fig. 3e) and with processes that surround synapses (Fig. 3g).

Macrophage-mediated phagocytosis in the periphery often requires antibody and complement deposition, however, no differences in the amount of endogenous mouse IgG coating VGLUT1-positive or

**Figure 3 | West Nile virus causes a loss in hippocampal CA3 synaptic terminals in mice and humans.** **a**, Immunostaining and quantification of colocalized presynaptic and postsynaptic puncta using the markers synaptophysin and Homer1, respectively, at 7 dpi in wild-type mice (mock ( $n = 7$ ) and WNV-NS5-E218A ( $n = 9$ ) mice). Data are the mean of 2 staining experiments.

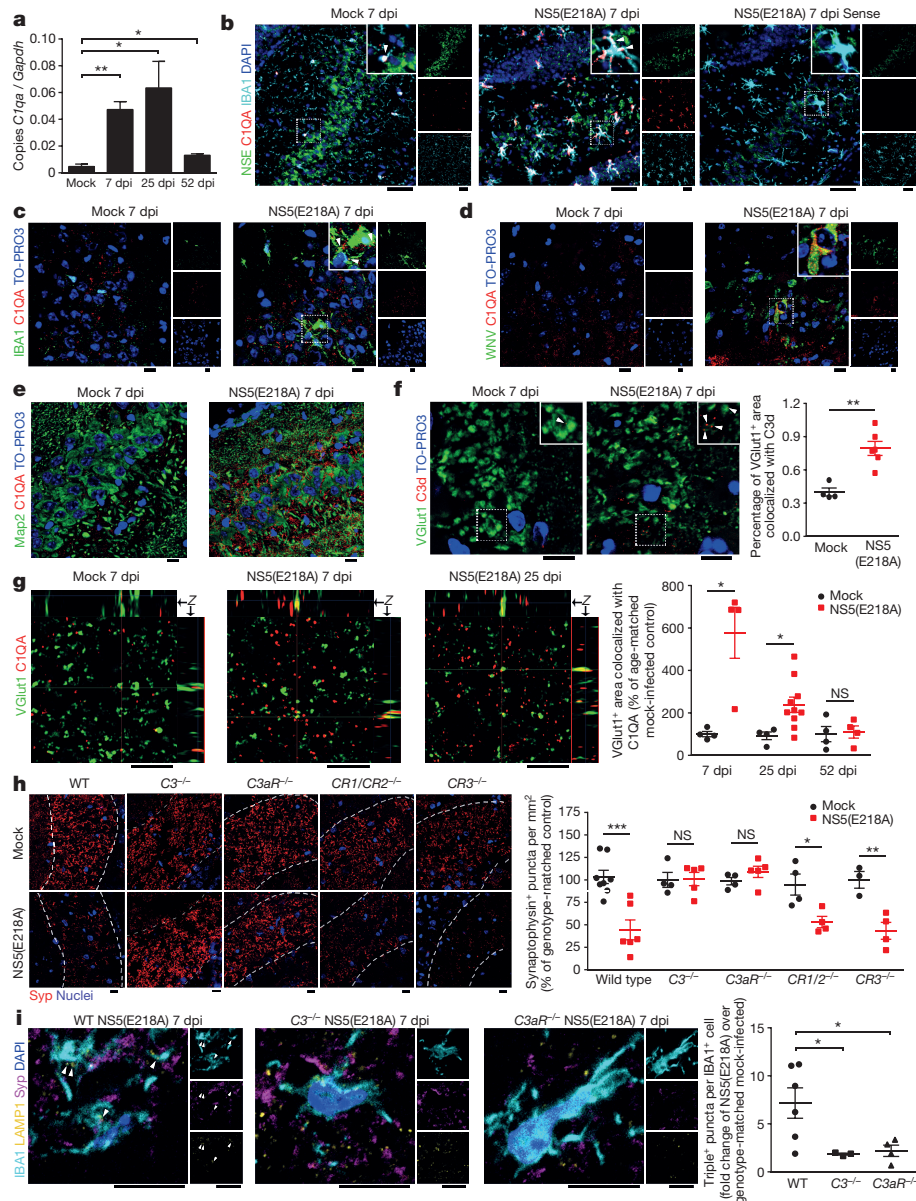
**b**, Immunostaining and quantification of staining area for glutamatergic presynaptic marker, VGlut1, at 25 dpi in mock or WNV-NS5-E218A-recovered mice with good or poor spatial memory performance (mock ( $n = 3$ ), WNV good ( $n = 5$ ), and WNV poor ( $n = 3$ ) mice). **c**, Immunostaining and quantification of synaptophysin<sup>+</sup> area in acute WNV encephalitis patients with age- and sex-matched controls. **d**, CX<sub>3</sub>CR1-GFP heterozygous mice were immunostained for synaptophysin and GFP with arrowheads depicting colocalization. Images shown are representative of 3 mice per group. **e**, Synaptic terminal elimination at 7 dpi is absent in WNV-NS5-E218A-infected *Il34*<sup>-/-</sup> mice, which have reduced numbers of microglia. **f**, Immunostaining showing the number of colocalized synaptophysin, LAMP-1 (lysosomal marker), and IBA1 puncta per IBA1<sup>+</sup> cell at 7, 25 and 52 dpi. **g**, Electron micrographs from mock or WNV-NS5-E218A-infected hippocampus at 7 dpi with immunohistochemical enhancement of IBA1. An IBA1<sup>+</sup> cell is shown surrounding a nearby synapse (right panels, boxed area, red arrowhead). Images shown are representative of 3 mice per group. All panels, \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ , NS, not significant, by two-tailed  $t$ -test. Error bars, s.e.m. Immunostaining, electron microscopy, and quantification performed within the hippocampal CA3 region. Scale bars, 10  $\mu$ m unless otherwise noted.



Homer1-positive, CA3 synaptic terminals at 7 and 25 dpi following WNV-NS5-E218A infection were observed compared to mock-infected controls (Extended Data Fig. 3f, g). Furthermore,  $\mu$ MT<sup>-/-</sup> mice, which lack mature B cells, eliminate synaptic terminals during acute WNV-NS5-E218A infection at similar levels as in wild-type mice (Extended Data Fig. 3h). Hippocampal upregulation of C1QA, the initiating factor of the classical complement cascade, was detected at 7, 25 and 52 dpi compared to mock-infected controls (Fig. 4a). Hippocampal *C1qa* mRNA in conjunction with IBA1 expression

was detected in mock-infected and WNV-NS5-E218A-infected mice at 7 dpi, the latter of which was increased (Fig. 4b). C1QA protein was detected in IBA1<sup>+</sup> cellular processes adjacent to or surrounding neurons (Fig. 4b, c).

C1QA also colocalized within WNV antigen-positive cells with neuronal morphology (Fig. 4d) and with Map2-positive neurites (Fig. 4e). Increased levels of the complement C3 cleavage product, C3d, also colocalized with VGlut1-positive terminals at 7 dpi (Fig. 4f). Levels of colocalization of C1QA with VGlut1-positive



**Figure 4 | Classical complement activation in neurons and microglia drives WNV-mediated synaptic terminal elimination.** **a**, qPCR analysis of hippocampal *C1qa* mRNA normalized to *Gapdh* in mock ( $n = 8$ ) or 7 dpi ( $n = 3$ ), 25 dpi ( $n = 8$ ) or 52 dpi ( $n = 7$ ) after WNV-NS5-E218A infection. **b**, Fluorescent *in situ* hybridization using RNA probes for neuron specific enolase (NSE) with sense or antisense C1QA coupled with immunostaining for IBA1 in WNV-NS5-E218A-infected or control mice with high magnification insets. Images are representative of 3 mice per group. Scale bars, 50  $\mu$ m. **c**, Immunostaining for C1QA protein and IBA1 with high magnification insets. Arrowheads depict colocalization. **d**, Immunostaining for C1QA protein and WNV antigen at 7 dpi with a WNV-infected neuron shown in high magnification inset. **e**, Immunostaining for C1QA with neuronal marker, Map2. **f**, Immunostaining shows colocalization of presynaptic marker,

presynaptic terminals were significantly increased in WNV-NS5-E218A-recovered animals compared with mock-infected controls at 7 and 25, but not 52 dpi (Fig. 4g). WNV-NS5-E218A-infected complement C3<sup>-/-</sup> and C3aR1<sup>-/-</sup>, but not CR1/2<sup>-/-</sup> or *Itgam*<sup>-/-</sup> (*Itgam* is also known as Cr3), mice showed no differences in synaptophysin-positive synaptic terminals compared to mock-infected controls (Fig. 4h). Presynaptic puncta within microglial (IBA1<sup>+</sup>) lysosomes were also decreased in WNV-NS5-E218A-infected complement C3<sup>-/-</sup> and C3aR1<sup>-/-</sup> compared to wild-type mice (Fig. 4i),

VGlut1, with C3d at 7 dpi (mock ( $n = 4$ ), WNV-NS5-E218A ( $n = 6$ )). **g**, Immunostaining showing colocalization of C1QA and VGlut1 with representative super-resolution micrographs shown. Scale bars, 5  $\mu$ m. **h**, Synaptophysin (Syp) immunostaining in WNV-NS5-E218A-infected wild-type, complement C3-null, complement receptors C3aR-null, CR1/2-null, and CR3-null mice at 7 dpi, normalized to age and genotype-matched, mock-infected controls. **i**, Immunostaining and quantification of number of colocalized (arrowheads) synaptophysin<sup>+</sup>, LAMP-1<sup>+</sup> (lysosomal marker), and IBA1<sup>+</sup> puncta per IBA1<sup>+</sup> cell in WNV-NS5-E218A-infected wild-type, C3-null and C3aR-null mice at 7 dpi (fold-change over control). All panels, \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ , NS, not significant, by two-tailed *t*-test. Error bars, s.e.m. Immunostaining and quantification performed within the hippocampal CA3 region. Scale bars, 10  $\mu$ m unless otherwise noted.

indicating that WNV-mediated elimination of synaptic terminals requires C3 and C3aR.

Many pathogens and pathogen-associated molecular patterns induce complement activation within the central nervous system, including WNV<sup>24</sup>, HIV<sup>25</sup>, and amyloid plaques<sup>26</sup>. Our study suggests that complement C3 and C3aR mediate presynaptic terminal loss in the hippocampi of mice that exhibit spatial learning defects during recovery from West Nile neuroinvasive disease. Microglia and recognition of C3 cleavage products by complement receptor C3aR, which is expressed by

both neurons and microglia<sup>27</sup>, are required for this process. Although astrocytes did not exhibit increased colocalization with synaptic terminals, we cannot completely rule out their contribution to this process, as their rates of lysosomal digestion differs from microglia<sup>28</sup>. Furthermore, alterations to NMDA receptor-mediated long-term potentiation could also contribute to altered synapse homeostasis and memory as several genes in this pathway are differentially expressed in WNV-poor learners (Supplementary Table 4). It is unknown whether complement labelled and eliminated terminals are connected to WNV infected or healthy neurons. In the context of neurotropic viral infection, elimination of presynaptic terminals may prevent trans-synaptic viral spread or aberrant signalling of infected neurons. Further studies will determine if complement- and microglial-dependent synapse elimination prevents neuron-to-neuron spread of WNV and other neurotropic viruses.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 August 2015; accepted 25 April 2016.**

- Sejvar, J. J. *et al.* Neurologic manifestations and outcome of West Nile virus infection. *J. Am. Med. Assoc.* **290**, 511–515 (2003).
- Klee, A. L. *et al.* Long-term prognosis for clinical West Nile virus infection. *Emerg. Infect. Dis.* **10**, 1405–1411 (2004).
- Petersen, L. R. *et al.* Estimated cumulative incidence of West Nile virus infection in US adults, 1999–2010. *Epidemiol. Infect.* **141**, 591–595 (2013).
- Stevens, B. *et al.* The classical complement cascade mediates CNS synapse elimination. *Cell* **131**, 1164–1178 (2007).
- Schafer, D. P. *et al.* Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron* **74**, 691–705 (2012).
- Daffis, S. *et al.* 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468**, 452–456 (2010).
- Szretter, K. J. *et al.* 2'-O methylation of the viral mRNA cap by West Nile virus evades Ifit1-dependent and -independent mechanisms of host restriction *in vivo*. *PLoS Pathog.* **8**, e1002698 (2012).
- Armah, H. B. *et al.* Systemic distribution of West Nile virus infection: postmortem immunohistochemical study of six cases. *Brain Pathol.* **17**, 354–362 (2007).
- Jarrard, L. E. On the role of the hippocampus in learning and memory in the rat. *Behav. Neural Biol.* **60**, 9–26 (1993).
- Sadek, J. R. *et al.* Persistent neuropsychological impairment associated with West Nile virus infection. *J. Clin. Exp. Neuropsychol.* **32**, 81–87 (2010).
- Clarke, P. *et al.* Death receptor-mediated apoptotic signaling is activated in the brain following infection with West Nile virus in the absence of a peripheral immune response. *J. Virol.* **88**, 1080–1089 (2014).
- Samuel, M. A., Morrey, J. D. & Diamond, M. S. Caspase 3-dependent cell death of neurons contributes to the pathogenesis of West Nile virus encephalitis. *J. Virol.* **81**, 2614–2623 (2007).
- McCandless, E. E., Zhang, B., Diamond, M. S. & Klein, R. S. CXCR4 antagonism increases T cell trafficking in the central nervous system and improves survival from West Nile virus encephalitis. *Proc. Natl Acad. Sci. USA* **105**, 11270–11275 (2008).
- Durrant, D. M., Robinette, M. L. & Klein, R. S. IL-1R1 is required for dendritic cell-mediated T cell reactivation within the CNS during West Nile virus encephalitis. *J. Exp. Med.* **210**, 503–516 (2013).
- Shrestha, B., Zhang, B., Purtha, W. E., Klein, R. S. & Diamond, M. S. Tumor necrosis factor alpha protects against lethal West Nile virus infection by promoting trafficking of mononuclear leukocytes into the central nervous system. *J. Virol.* **82**, 8956–8964 (2008).
- Habjan, M. *et al.* Sequestration by IFIT1 impairs translation of 2'-O-unmethylated capped RNA. *PLoS Pathog.* **9**, e1003663 (2013).
- Barnes, C. A. Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. *J. Comp. Physiol. Psychol.* **93**, 74–104 (1979).
- Hickman, S. E. *et al.* The microglial sensome revealed by direct RNA sequencing. *Nat. Neurosci.* **16**, 1896–1905 (2013).
- Chu, Y. *et al.* Enhanced synaptic connectivity and epilepsy in C1q knockout mice. *Proc. Natl Acad. Sci. USA* **107**, 7975–7980 (2010).
- Mehlhop, E. & Diamond, M. S. Protective immune responses against West Nile virus are primed by distinct complement activation pathways. *J. Exp. Med.* **203**, 1371–1381 (2006).
- Clarke, P., Leser, J. S., Bowen, R. A. & Tyler, K. L. Virus-induced transcriptional changes in the brain include the differential expression of genes associated with interferon, apoptosis, interleukin 17 receptor A, and glutamate signaling as well as flavivirus-specific upregulation of tRNA synthetases. *MBio* **5**, e00902–e00914 (2014).
- Ménard, C. *et al.* Glutamate presynaptic vesicular transporter and postsynaptic receptor levels correlate with spatial memory status in aging rat models. *Neurobiol. Aging* **36**, 1471–1482 (2015).
- Wang, Y. *et al.* IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. *Nat. Immunol.* **13**, 753–760 (2012).
- Mehlhop, E. *et al.* Complement activation is required for induction of a protective antibody response against West Nile virus infection. *J. Virol.* **79**, 7466–7477 (2005).
- Ebenbichler, C. F. *et al.* Human immunodeficiency virus type 1 activates the classical pathway of complement by direct C1 binding through specific sites in the transmembrane glycoprotein gp41. *J. Exp. Med.* **174**, 1417–1424 (1991).
- Veerhuis, R. *et al.* Cytokines associated with amyloid plaques in Alzheimer's disease brain stimulate human glial and neuronal cell cultures to secrete early complement proteins, but not C1-inhibitor. *Exp. Neurol.* **160**, 289–299 (1999).
- Lian, H. *et al.* Astrocyte-microglia cross talk through complement activation modulates amyloid pathology in mouse models of Alzheimer's disease. *J. Neurosci.* **36**, 577–589 (2016).
- Chung, W.-S. *et al.* Astrocytes mediate synapse elimination through MEGF10 and MERTK pathways. *Nature* **504**, 394–400 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** Funding for this research was provided by the NIH F31 NS077640 (M.J.V.), R01 NS052632 (R.S.K.), and U19 AI083019 (R.S.K. and M.S.S.). The authors would like to thank J. Atkinson and X. Wu for reagents and M. Diamond for critical reading of the manuscript.

**Author Contributions** M.J.V. and R.S.K. contributed to the study design. M.J.V., C.G., D.D., D.M.D., B.B., A.S., J.Y., C.P.-T., A.F., D.K.W., K.F., X.J., S.M., J.K.R., J.R.G., R.E.S., B.S. and R.S.K. contributed to data collection and/or interpretation. C.G., J.R.B. and M.S.S. developed single-strand PCR assays for WNV. B.K.D., K.L.T. identified, collected and provided patient samples. M.J.V. and R.S.K. wrote the paper. All authors discussed and commented on the manuscript.

**Author Information** Microarray data has been deposited in the Gene Expression Omnibus under the accession number GSE72139. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S.K. (rklein@dom.wustl.edu).

## METHODS

**Animals.** At the outset of all experiments, 8–10-week-old male and female mice were used. C57BL/6J and *CD11b*<sup>−/−</sup> (*CR3*<sup>−/−</sup>) mice were obtained from Jackson Laboratories. *C3*<sup>−/−</sup>, *CR1/2*<sup>−/−</sup> (in mice, CR1 and CR2 are splice variants both derived from the mouse *Cr2* gene), and *C3aR*<sup>−/−</sup> mice (>10 generations backcrossed to C57BL/6J) were obtained from John Atkinson (Washington University), CX<sub>3</sub>CR1–GFP<sup>+/−</sup> mice (>10 generations backcrossed to C57BL/6J) were obtained from Richard Ransohoff (Lerner Research Institute, Cleveland Clinic Foundation),  $\mu$ MT<sup>−/−</sup> mice (C57BL/6 background) were obtained from Michael Diamond (Washington University), and *IL34*<sup>−/−</sup> mice (C57BL/6 background) were obtained from Marco Colonna (Washington University). All mice were randomly assigned to control or experimental groups at the beginning of each experiment. All experimental protocols were performed in compliance with the Washington University School of Medicine Animal Studies Committee (protocol number 20140122).

**Mouse models of WNV infection.** The WNV strain 3000.0259 was isolated in New York and passaged once in C6/36 *Aedes albopictus* cells to generate an insect-cell-derived stock. Then 100 plaque-forming units (pfu) of WNV-NY99 were delivered in 50  $\mu$ l to the footpad of anaesthetized mice. WNV-NS5-E218A, which harbours a single point mutation in the 2' O-methyl-transferase gene, was obtained from Michael Diamond (Washington University) and passaged in Vero cells as described previously<sup>6</sup>. Deeply anaesthetized mice were administered with 10<sup>4</sup> pfu of WNV-NS5-E218A or 10 pfu of WNV-NY99 in 10  $\mu$ l into the brain's third ventricle via a guided 29 gauge needle.

Stock titres of all viruses were determined by using BHK21 cells for viral plaque assay as previously described<sup>14</sup>.

**Leukocyte isolation and flow cytometry.** Cells were isolated from brains of wild-type mice at day 7 post-infection and stained with fluorescently conjugated antibodies to CD4, CD8, CD11b, and CD45 as previously described<sup>14</sup>. Data collection and analysis were performed with an LSRII flow cytometer using FlowJo software.

**Behavioural testing. Test for anxiety and locomotor behaviour.** The open-field test was used to assess baseline differences in anxiety or locomotor behaviour, before Barnes maze experiments. A standard Open Field arena (54 × 54 cm, custom built) was used, consisting of a simple square box with a grid (6 squares per side) along the base. Animals were placed in the centre of the arena and allowed to explore for 5 min. The arena was decontaminated with 70% ethanol between each trial. Locomotor activity was assessed by counting the number of lines the animal crossed during the testing period, and anxiety was assessed by counting the number of times the animal crossed through the centre of the field. Behaviour was recorded using a camera (Canon PowerShot SD1100 IS), and a blinded experimenter scored the trial. Any mice that jumped out of the open-field maze were excluded from analyses.

**Test for visual spatial memory.** The Barnes maze was used to assess visual spatial memory. An elevated Barnes maze (91.4 cm diameter, custom built) containing 19 empty holes and 1 target hole with a hidden escape chamber was used for testing (5 cm diameter holes were evenly spaced around the table, 6.35 cm from the edge). Visual cues were placed around the room and remained in the same location during the entire testing period. Mice were tested on the Barnes maze over the course of 5 consecutive days. Each mouse received two trials per day, spaced exactly 30 min apart. For each trial, the mouse was placed in the centre of the maze in a covered start box for 10 s, and removal of the box signalled the start of a trial. Each mouse was given 3 min to explore the maze and find the target hole. Mice that did not enter the target hole within 3 min were gently guided into the hole. After each trial, the mouse remained in the target hole for exactly 1 min, and then was returned to its home cage. The maze was decontaminated with 70% ethanol between each trial. The numbers of errors (nose pokes over non-target holes) and the latency to find the target hole (amount of time elapsed before nose poke over target hole) were measured. Behaviour was recorded using a camera (Canon Powershot SD1100IS), and a blinded experimenter scored the trials. Any mice which fell off the Barnes maze table during any trial were excluded from analyses. No randomization was required for these studies.

**Immunohistochemistry.** Following perfusion with ice-cold PBS and 4% PFA, brains were immersion-fixed overnight in 4% PFA, followed by cryoprotection in two exchanges of 30% sucrose for 72 h, then frozen in OCT (Fisher). 9  $\mu$ m-thick fixed-frozen coronal brain sections were washed with PBS and permeabilized with 0.1% Triton X-100 (Sigma-Aldrich), and nonspecific antibody was blocked with 5–10% normal goat serum (Santa Cruz Biotechnology) for 1 h at room temperature. Mouse on mouse kit (MOM basic kit, Vector) was used as per the manufacturer's protocol when detecting synaptophysin to reduce endogenous mouse antibody staining. After block, slides were exposed to primary antibody or isotype matched IgG overnight at 4°C, washed with 1 × PBS and incubated with secondary antibodies for 1 h at room temperature. Nuclei were counterstained with TO-PRO-3 (Invitrogen) and coverslips were applied with vectashield (Vector).

Immunofluorescence was analysed using a Zeiss LSM 510 laser-scanning confocal microscope and accompanying software (Zeiss). Positive immunofluorescent signals were quantified by a blinded experimenter using the NIH Image analysis software, ImageJ.

TUNEL staining was performed using the TMR-red *in situ* cell death detection kit (Roche) as per manufacturer's instructions. C1QA staining was performed as previously described<sup>29</sup>.

**Antibodies.** C1QA (undiluted, described previously<sup>29</sup>), WNV (1:100, described previously<sup>14</sup>), rat anti-GFAP (1:200, Invitrogen catalogue number 13-0300), rabbit anti-IBA1 (1:100, WAKO catalogue number 019-19741), mouse anti-NeuN-biotin (1:100, Millipore catalogue number MAB 377B), chicken anti-GFP (1:1,000, Abcam catalogue number 13970), rabbit anti-Synapsin1 (1:200, Millipore catalogue number Ab1543), mouse anti-synaptophysin (1:50, DAKO catalogue number M0776 or 1:50 Abcam catalogue number ab8049), guinea-pig anti-VGlu1 (1:300, Synaptic Systems catalogue number 135304), rabbit anti-Homer1 (1:200, Synaptic Systems 160003), rabbit anti-C3d (1:500, DAKO catalogue number A0063), rabbit anti S100 $\beta$  (1:300, Abcam catalogue number ab52642), rat anti-Lamp1 (1:50, BD Pharmingen catalogue number 553792), and rat anti-CD68 (1:200, Serotec catalogue number MCA1957).

Secondary antibodies conjugated to Alexa-488, Alexa-555, or Alexa-633 (Invitrogen) were used at 1:400 dilution.

**MRI.** Mice were intracardially perfused, first with ice-cold PBS and then with a mixture of 4% PFA and 10% Multihance (gadobenate dimeglumine, Bracco Diagnostics, Princeton, NJ). Heads were further fixed in 4% PFA for 24 h before being trimmed of extraneous tissue around the skull (to minimize the field of view). Heads were then placed in 1% Multihance in PBS until being imaged several days later. *Ex vivo*, whole-head MR imaging experiments were performed at 4.7 T using an Agilent/Varian (Santa Clara, CA) DirectDrive1 small-animal scanner. Data were collected with a custom-made RF foil coil that fits tightly around the head using a 3D, T1-weighted gradient echo sequence with the following parameters: TR = 105 ms, TE = 6 ms, flip angle = 90°, isotropic resolution = (0.0625 mm)<sup>3</sup>, and scan time  $\sim$ 11 h. Regions of interest were manually drawn for the whole brain and hippocampus in ITK-SNAP (<http://www.itksnap.org>) from which volumes were calculated.

**Collection and purification of hippocampal RNA.** Mice were perfused with ice-cold PBS then hippocampi were dissected and snap-frozen in Tri-reagent (Ambion). Hippocampal tissue was then homogenized, and RNA purified as previously described<sup>30</sup> using the RNA Ribopure kit (Ambion). RNA was precipitated with 25 mM ammonium acetate in 100% ethanol at −80°C overnight, resuspended in RNase-free H<sub>2</sub>O, and checked for purity. RNA was then treated with RNase-out and DNase I (Invitrogen) as per the manufacturer's protocol.

**Microarray.** Hippocampal RNA was isolated as described above and submitted to the Washington University Genome Technology Access Center. The total RNA quality and quantity were then determined by Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA) and NanoDrop ND-1000 Spectrophotometer (Thermo Scientific NanoDrop, Wilmington, DE), according to manufacturer's recommendations, respectively. A total of 400 ng of RNA transcripts from each sample were amplified by T7 linear amplification with the MessageAmp TotalPrep Amplification kit (Life Technologies-Ambion, Austin, TX). Then 1.5  $\mu$ g of each amplified and biotinylated RNA (aRNA) sample was hybridized onto Illumina MouseWG-6 v2 expression beadchips, followed by cy3 streptavidin-based staining, washing, and scanning, according to Illumina standard protocol. The iScan scanner-created image data were loaded into Illumina GenomeStudio (v2011) for generation of expression values and data normalizations. Only those probes that were detected at  $P < 0.05$  in at least one of the samples were used in downstream statistical analysis. Background subtracted and quantile normalized data were used in statistical analysis for identification of differentially expressed genes with one-way ANOVA test using Partek Genomics Suite (v6.6, St. Louis, MO). All original  $P$  values in the ANOVA analysis were adjusted by  $q$ -value based multiple test correction<sup>31</sup>. KEGG pathway analysis was performed using DAVID bioinformatics database functional annotation tool v6.7 (ref. 32). Microarray data has been deposited in the Gene Expression Omnibus (accession number GSE72139).

**Real-time quantitative RT-PCR.** cDNA was synthesized using random hexamers, oligodT15, and MultiScribe reverse transcriptase (Applied Biosystems). A single reverse transcription master mix was used to reverse transcribe all samples to minimize differences in reverse transcription efficiency. The following conditions were used for reverse transcription reactions: 25°C for 10 min, 48°C for 30 min, and 95°C for 5 min.

For all primer sets except for the strand-specific WNV PCR reaction (detailed below), PCR reactions were prepared using Power SYBR Green PCR mastermix and calculated copies were normalized against copies of the housekeeping gene, *Gapdh*.

Primers are listed in Supplementary Table 6.

**WNV strand specific real-time RT-PCR.** Two-step strand specific RT-PCR was performed using GVA and T7 tagged primers during cDNA synthesis as a modification to a procedure previously described<sup>33</sup>. First, cDNA was synthesized using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) per the manufacturer's instruction, with the addition of 1 pmol of the positive-strand primer GVA\_TxsspE1229 and 1 pmol of the negative strand primer T7\_TxsspE1160 to the 10 µl reaction mixture. Strand specific cDNA was then amplified in a qPCR reaction using primers directed against the strand specific WNV sequence and the tag sequence. Each 12.5 µl qPCR reaction mixture contained 6.25 µl TaqMan Gene Expression Master mix, 10 pmol tag primer (positive mix: GVA, negative mix: T7), 10 pmol strand specific primer (positive mix: TXsspE1160, negative mix: TXsspE1229), 2.5 pmol strand specific probe (WNVssPEProbe) and 50 ng strand specific cDNA. Thermal cycling was performed using Applied Biosystems ViiA 7 Real-Time PCR system with a 384-well block. Copies were calculated based on a standard curve generated from purified PCR product from positive strand or negative strand reactions, and normalized to amount of *Gapdh* in each sample.

For the following sequences, the bold sequence represents the tag sequence, while the underlined sequence represents the strand specific sequence: **GVA\_TxsspE1229:** 5'-TTTGCTAGCTTTAGGACCTACTATATCTACCTGGTCAGCACGTTTGTCTATTG-3'; **T7\_TxsspE1160:** 5'-GCGTAATACGACTCCTATATCAGCGATCTCTCCACCAAAG-3'; **TxsspE1229:** 5'-GGGTACGACGTTTGTCTATTG-3'; **TxsspE1160:** 5'-TCAGCGATCTCTCCACCAAAG-3'; **T7tag:** 5'-GCGTAATACGACTCACTATA-3'; **GVAtag:** 5'-TTTGCTAGCTTTAGGACCTACTATATCTACCT-3'; **WNVssPEProbe:** FAM-TGCCCGACCATGGGAGAAGCTC-TAMRA.

**Quantification of synaptic terminals.** ImageJ was used to threshold single-plane confocal images, draw a region of interest encompassing the CA3 mossy fibres, and to quantify the number of synaptophysin or VGlut1<sup>+</sup> puncta containing between 0.5 and 25 square µm<sup>2</sup> in area. For each mouse, at least 12 images at 63× magnification were counted, which were derived from at least 4 fixed-frozen coronal sections spaced 50 µm apart.

**Quantification of synaptic terminal engulfment by IBA1<sup>+</sup> cells.** ImageJ was used to threshold single-plane confocal images, draw a region of interest encompassing each IBA1<sup>+</sup> cell within the hippocampal CA3, and to quantify the number of synaptophysin<sup>+</sup>, Lamp1<sup>+</sup>, IBA1<sup>+</sup> puncta between 0.2 and 25 µm<sup>2</sup> in area within each IBA1<sup>+</sup> cell. For each mouse, at least 8 images at 63× magnification were counted, which were derived from at least 4 fixed-frozen coronal sections spaced 50 µm apart.

**Electron microscopy.** Mice were perfused with 4% PFA, the brain was removed, and immersion fixed for 24 h at room temperature. Tissue was washed in PBS, and 100 µm sections were cut from regions of interest using a vibratome. Sections were incubated in 0.5% gelatin, 5% horse serum, and 0.01% saponin in PBS for 5 h on a rotator at room temperature. Sections were then incubated for 48 h at 4 °C on a rotator with rabbit anti-IBA1 (1:600; Wako). After washing in dPBS, sections were incubated overnight at 4 °C on rotator with donkey anti-rabbit biotinylated secondary antibody, (1:500; Rockland) in 0.5% gelatin and 5% horse serum in dPBS. Sections were again washed, and then incubated with streptavidin-HRP (1:1000; Rockland, S000-03) for 3 h at room temperature, followed by another wash. HRP was visualized using the DAB Substrate Kit (Cell Marque, 957D) for 5 min, washed and then fixed with 2% PFA, 2.5% glutaraldehyde, in PBS for 30 min, followed by a wash in PBS. Sections were post-fixed in 1% osmium tetroxide in PBS for 30 min at room temperature, washed in PBS, and then dehydrated in sequential concentrations of ethanol for 30 min each. Sections were then infiltrated with 1:1 Spurr's resin and 100% ethanol overnight on a rotator, followed by two changes in pure Spurr's resin over 24 h. Sections were embedded with Aclar film (EMS, 50425), and polymerized at 60 °C for 48 h. The hippocampus was trimmed from the polymerized section, glued to a previously prepared block of Spurr's resin, and allowed to cure for at least 24 h. Using a Diatome ultra 45° diamond knife and a LEICA Ultracut UC7, blocks were sectioned at 500 nm to confirm location of the tissue and the positivity of antibodies. Once confirmed, 90 nm sections were cut and picked up onto 200 hex mesh, formvar-carbon coated copper grids (Ted Pella, 01800-F).

Images were digitally captured using a JEOL 1200 EX II transmission electron microscope with AMT digital camera.

**Three-dimensional reconstruction of confocal z-stack images.** Confocal z-stack images taken with a Zeiss LSM 510 META microscope at 63× magnification, consisting of at least 10 images were transformed into 3D reconstruction videos using Volocity 3D image analysis software (PerkinElmer).

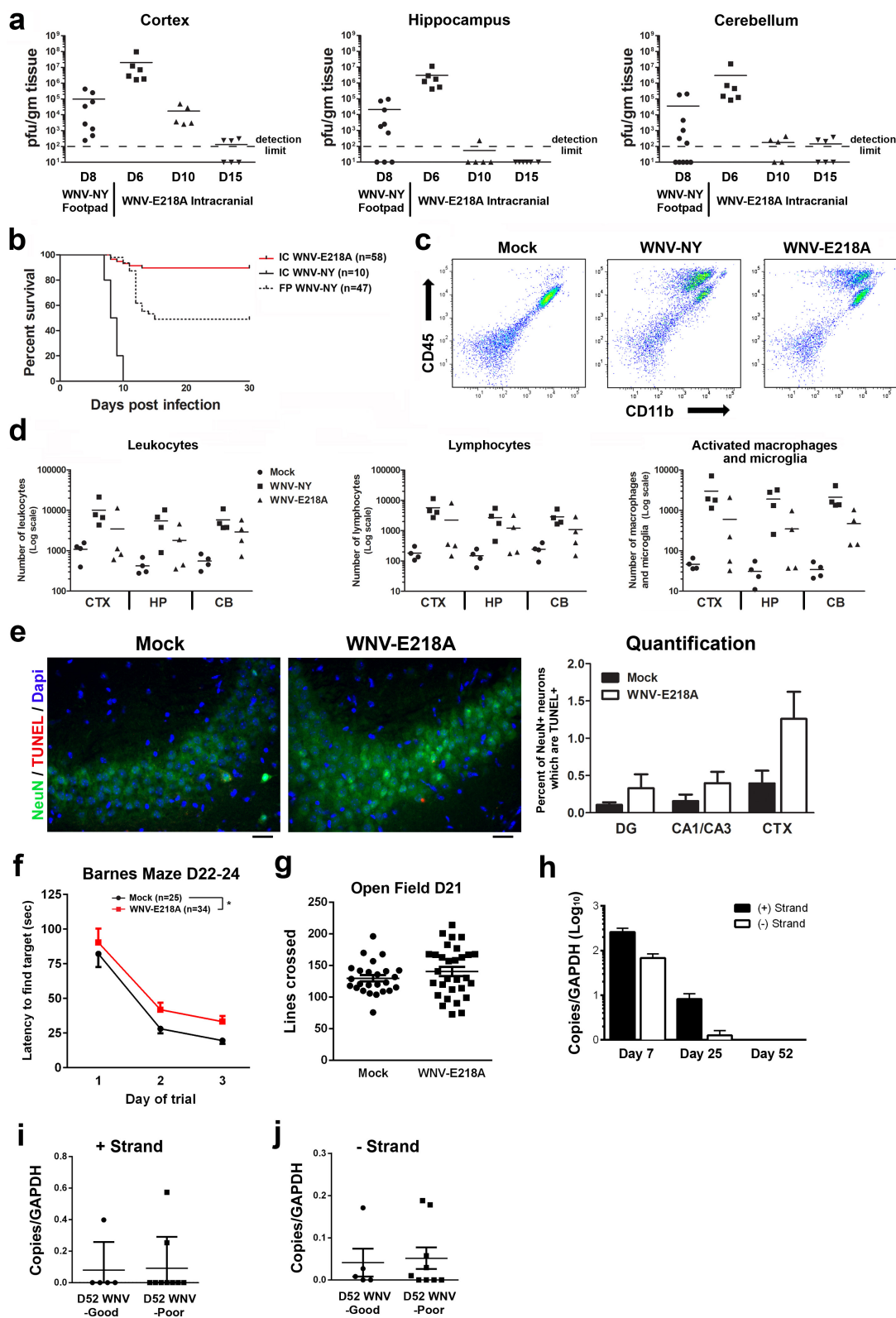
**Super-resolution microscopy of presynaptic terminals.** Sections were imaged using a Zeiss ELYRA PS1 microscope. 0.101 µm optical slices of the CA3 region of the hippocampus were captured and images subsequently processed using Zeiss SIM algorithms to generate structured illumination images. Zeiss Zen software was used to generate orthogonal viewpoints showing colocalization of VGlut1 and C1q in the x, y and z planes.

**Human tissue.** Human autopsy hippocampal tissue embedded in paraffin was obtained from St. Louis University Medical Center (St. Louis, MO) and Presbyterian / St. Luke's Medical Center (Denver, CO). Sections were first deparaffinized and then boiled in 10 mM sodium citrate buffer for 30 min, for antigen retrieval before staining with anti-synaptophysin antibody (DAKO, 1:50).

**In situ hybridization.** Fluorescent *in situ* hybridization was performed on 9 µm coronal brain sections that were 4% PFA-fixed and frozen. C1QA and NSE were used as double mRNA staining and IBA1 as immune-staining. C1QA and NSE anti-sense RNA was made and labelled with cyanine and fluorescein, respectively, using an RNA labelling kit (Roche). C1QA was amplified by PCR from pCMV SPORT6 C1QA plasmid (Openbiosystem MMM1013-63584), using forward 5'-GGCATCCGGACTGGTATCCGAGG-3' and reverse 5'-GGTAAATGCGACCCCTTTCGCGGG-3' primers, which was digested with SalI and transcribed with T7 promoter. The RNA probes were incubated overnight at 64 °C, and then detected with antibodies against fluorescein and cyanine (Roche). The staining reaction was then amplified with a TSA staining kit (PerkinElmer). A rabbit anti-IBA1 (Wako) antibody was used to label macrophages and microglia and detected with a donkey anti-rabbit 647 antibody (Life technologies). Stained sections were then imaged and analysed on a Zeiss AX10 fluorescent microscope.

**Statistical analysis.** To determine mouse group sizes for virological or immunological studies, power analysis was performed using the following values: probability of type I error = 0.05, power = 80%, fivefold hypothetical difference in mean, and population variance of 25-fold (virological studies) or 12-fold (immunological studies). Results from Barnes maze spatial learning and memory studies were compared by repeated measures two-way ANOVA. Microarray data was analysed by one-way ANOVA and fold change greater than 1.5, false discovery rate  $q < 0.05$  to correct for multiple hypotheses for mock vs WNV-all comparison,  $P < 0.05$  for WNV good learners vs WNV poor learners. Variance between groups was equivalent except for cases noted within figure legends and the comparison for number of presynaptic terminals in Fig. 3a and the comparisons at 7 and 25 dpi in Fig. 3f, in which a Welch's correction on the two-tailed t-test was used to correct for unequal variance. All other experiments were compared by Student's two-tailed t-test, with  $*P < 0.05$  considered significant. Power calculations using results observed in a pilot study in which WNV-recovered mice exhibit a two-fold increase in peak errors compared with mock-infected animals indicate that at least 15 recovered mice per group will be required to obtain statistical significance ( $P < 0.05$ ) on Barnes maze testing.

29. Stephan, A. H. et al. A dramatic increase of C1q protein in the CNS during normal aging. *J. Neurosci.* **33**, 13460–13474 (2013).
30. Sun, T., Vasek, M. J. & Klein, R. S. Congenitally acquired persistent lymphocytic choriomeningitis viral infection reduces neuronal progenitor pools in the adult hippocampus and subventricular zone. *PLoS One* **9**, e96442 (2014).
31. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
32. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).
33. Samuel, M. A. & Diamond, M. S. Alpha/beta interferon protects against lethal West Nile virus infection by restricting cellular tropism and enhancing neuronal survival. *J. Virol.* **79**, 13350–13361 (2005).



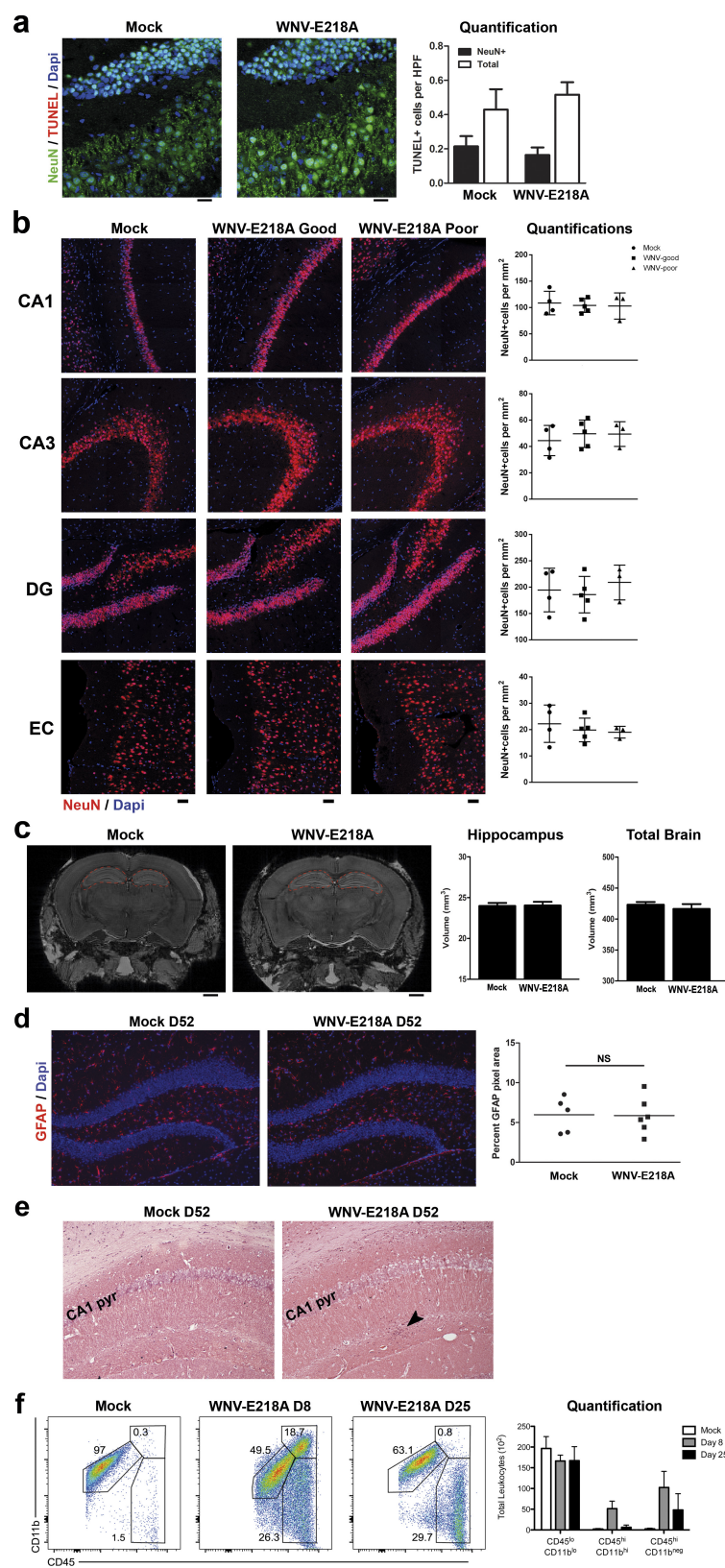
Extended Data Figure 1 | See next page for caption.

**Extended Data Figure 1 | Murine intracranial infection with attenuated WNV-NS5-E218A induces similar viral loads and inflammatory response as wild-type WNV-NY99, but greater overall survival.**

**a**, Plaque assay for infectious virus (measured in plaque-forming units per g of tissue) performed on dissected brain tissue at various days post-infection with either footpad infection with  $10^2$  pfu of WNV-NY99 or intracranial infection with  $10^4$  pfu of WNV-NS5-E218A. Each point represents an individual mouse. **b**, Survival curves of mice infected at 8-weeks-old by the footpad with WNV-NY99 or intracranially with WNV-NY99 or WNV-NS5-E218A. **c**, Flow cytometric analysis of dissected cortex, hippocampus and cerebellum at 6 dpi with WNV-NY99 and WNV-NS5-E218A with plots for CD45 and CD11b. **d**, Quantification of flow cytometry data from **c**. Shown are numbers of leukocytes (CD45<sup>high</sup>), lymphocytes (CD45<sup>high</sup>, CD11b<sup>low</sup>), and activated macrophages and microglia (CD45<sup>high</sup>, CD11b<sup>high</sup>) compared to mock-infected controls ( $n = 4$  mice per group). **e**, Immunostaining and counts for TUNEL staining for apoptotic cells with co-staining for the neuronal marker, NeuN, during peak infection (7 dpi) of WNV-NS5-E218A ( $n = 5$ )

compared to mock-infected controls ( $n = 4$ ). DG, dentate gyrus, CTX, entorhinal, perirhinal, and visual cortex. **f**, Some mice were tested at 22 dpi on a three-day version of the Barnes maze, and evaluated for latency to find target hole ( $*P < 0.05$  by repeated measures two-way ANOVA).

**g**, Prior to Barnes maze testing, mice were tested on open field for total lines crossed in 2 min at 21 dpi. **h**, qPCR for positive strand (non-replicating strand) and negative strand (replicating) WNV envelope protein message remaining in hippocampal tissue at 7, 25 and 52 dpi ( $n = 13, 4$ , and 14 mice per group for 7, 25, and 52 dpi, respectively), measured in copies per *Gapdh*. **i**, qPCR for positive strand WNV envelope protein at 52 dpi in WNV good learners (fewer than 8 errors on day 2 of Barnes maze,  $n = 5$ ) and WNV poor learners (greater than 9.5 errors on day 2 of Barnes maze,  $n = 9$ ). **j**, qPCR for negative strand WNV envelope protein at 52 dpi in WNV good learners (fewer than 8 errors on day 2 of Barnes maze,  $n = 5$ ) and WNV poor learners (greater than 9.5 errors on day 2 of Barnes maze,  $n = 9$ ). Result was not significant by Student's two-tailed *t*-test.

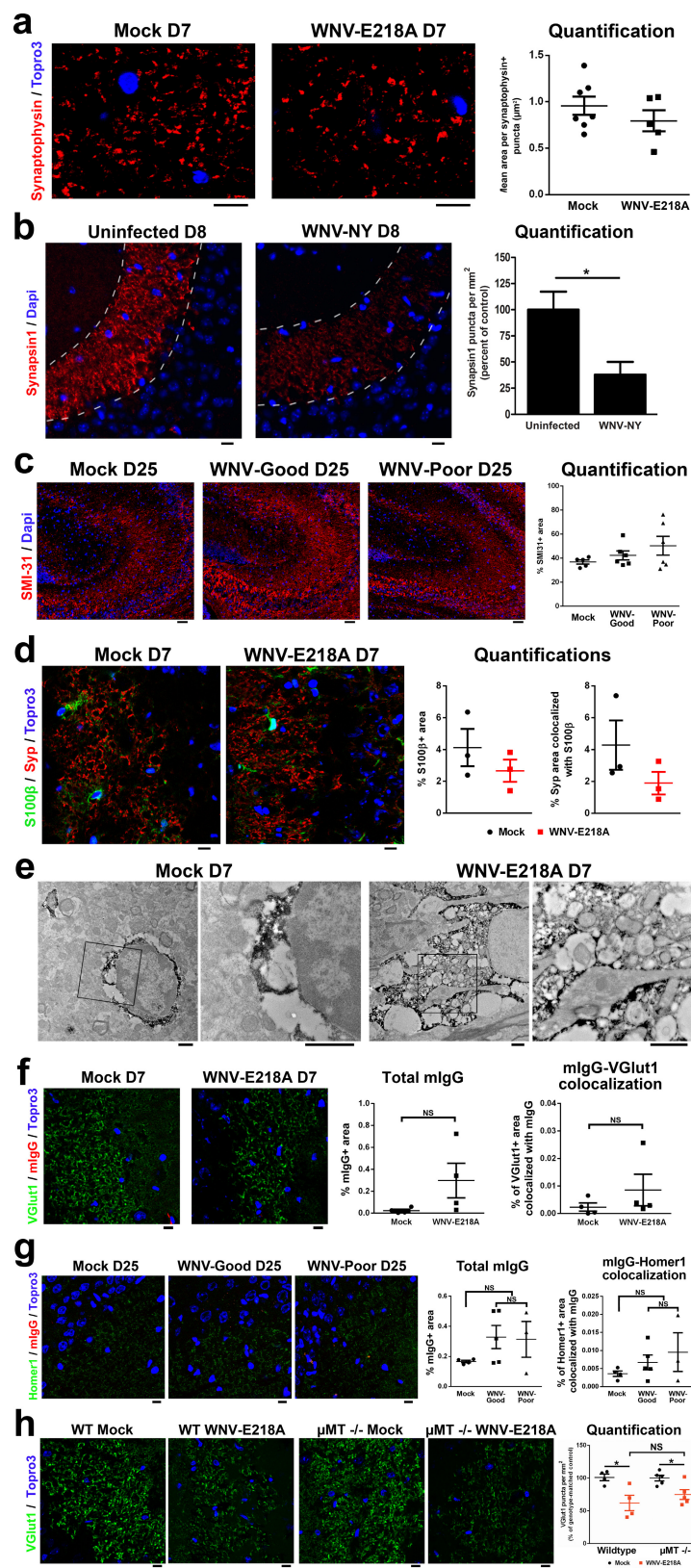


Extended Data Figure 2 | See next page for caption.

**Extended Data Figure 2 | At 25–52 days post-WNV-NS5-E218A infection, mice do not show any appreciable loss in brain volume, neuron or astrocyte numbers, or macrophage infiltration.**

**a**, Immunostaining for the neuronal marker, NeuN, with TUNEL staining for apoptotic cells within the hippocampus at 52 dpi. Quantification of the number of TUNEL<sup>+</sup> neurons and total TUNEL<sup>+</sup> cells is shown in mock ( $n = 3$ ) and WNV-NS5-E218A ( $n = 6$ ). Scale bar, 20  $\mu\text{m}$ . **b**, Immunostaining and quantification of the number of NeuN<sup>+</sup> neurons per  $\text{mm}^2$  within the CA1, CA3, dentate gyrus and entorhinal cortex at 25 days after mock ( $n = 4$ ) or WNV-NS5-E218A infection. WNV-infected animals were subdivided into good ( $n = 5$ ) and poor ( $n = 3$ ) learners. Scale bar, 100  $\mu\text{m}$ . **c**, Post-mortem mouse brains were imaged by MRI at 52 dpi to determine tissue volume of the hippocampus (outlined in red) and total brain ( $n = 5$  mice per group). Scale bar, 1 mm. Not significant by Student's

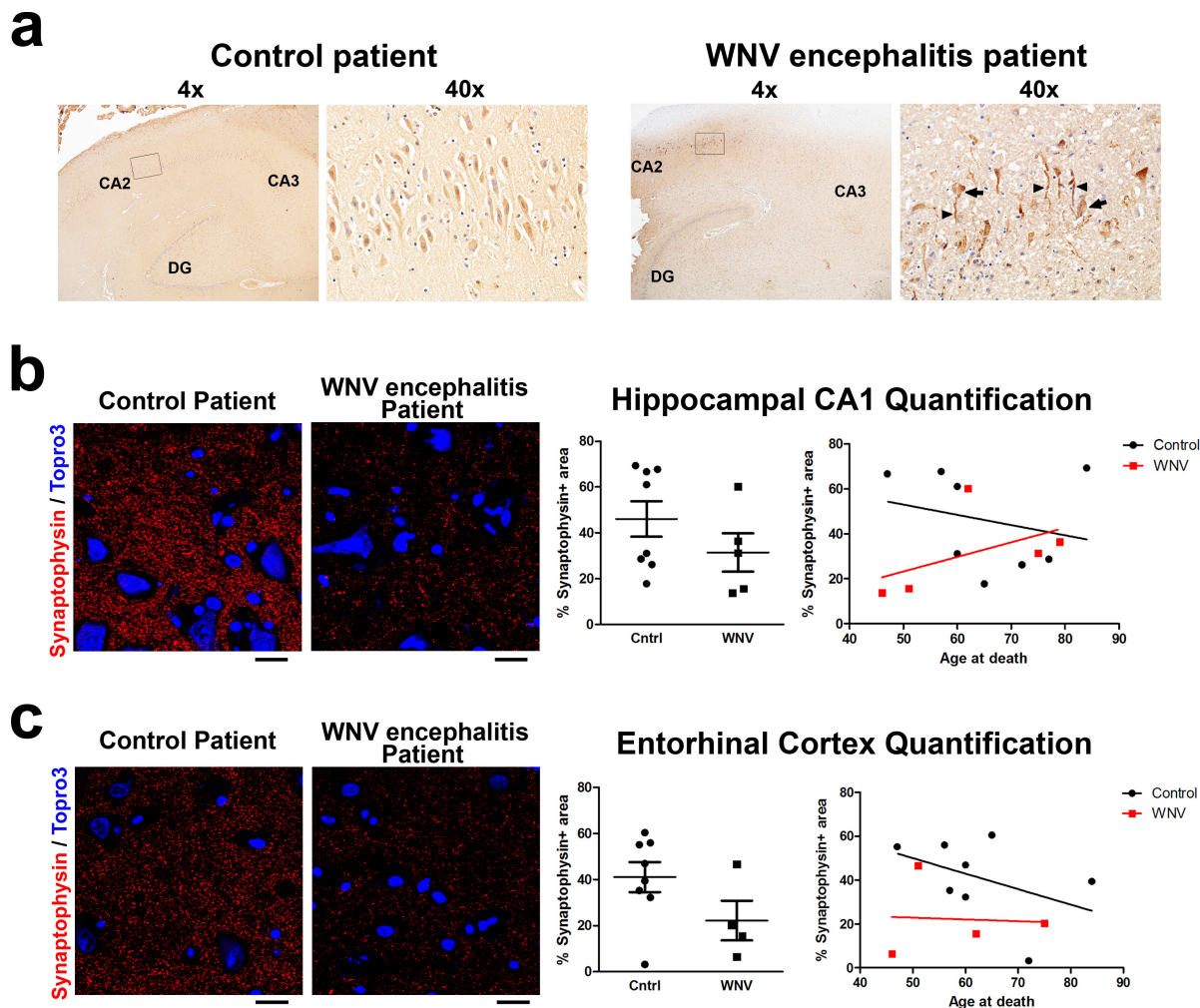
two-tailed  $t$ -test ( $P < 0.05$  considered significant). **d**, Immunostaining for the reactive astrocyte marker, GFAP, shows that WNV-NS5-E218A-infected mice do not exhibit greater hippocampal astrocyte activation than mock-infected controls at 52 dpi. NS, not significant by Student's two-tailed  $t$ -test. **e**, Haematoxylin and eosin (H&E) staining was performed at 52 dpi in WNV-NS5-E218A-recovered and mock-recovered mice. Occasional microglial nodules (arrowhead) surrounded by lymphocytes were observed within the hippocampus. CA1 pyr, CA1 pyramidal layer. **f**, Flow cytometric analysis of whole brain from mock and WNV-NS5-E218A-infected mice at 8 and 25 dpi was performed to determine numbers of microglia ( $\text{CD45}^{\text{low}}$ ,  $\text{CD11b}^{\text{low}}$ ), macrophages ( $\text{CD45}^{\text{high}}$ ,  $\text{CD11b}^{\text{high}}$ ), and lymphocytes ( $\text{CD45}^{\text{high}}$ ,  $\text{CD11b}^{\text{negative}}$ ). Note the decrease in macrophage population from 7 to 25 dpi.



Extended Data Figure 3 | See next page for caption.

**Extended Data Figure 3 | Despite synaptic terminal loss, no changes to synaptic terminal size, axons, or astrocyte or antibody association with terminals during WNV infection.** **a**, Immunostaining for the presynaptic marker, synaptophysin, at 7 dpi comparing mock ( $n = 7$ ) with WNV-NS5-E218A-infected ( $n = 5$ ) mice. Quantification of synaptophysin<sup>+</sup> puncta size was performed within the hippocampal CA3. Scale bar, 10  $\mu\text{m}$ . **b**, Immunostaining for the presynaptic marker, synapsin1, within the hippocampal CA3 in uninfected controls ( $n = 3$ ) and footpad-infected WNV-NY-1999 ( $n = 4$ ) at 8 dpi. Quantification was performed on the numbers of synapsin1<sup>+</sup> puncta per  $\text{mm}^2$  with  $*P < 0.05$  considered significant. **c**, Immunostaining within the hippocampal CA3 for SMI-31, which detect phosphorylated neurofilament and marks axons at 25 dpi ( $n = 5$ –6 mice per group). Quantification of the area of SMI-31 per  $\text{mm}^2$  (not significant by Student's  $t$ -test). **d**, Immunostaining within the hippocampal CA3 for the presynaptic marker, synaptophysin, co-labelled with the astrocyte marker, S100 $\beta$  at 7 dpi ( $n = 3$  mice per group). Quantification of the percentage of total S100 $\beta$ <sup>+</sup> area and synaptophysin<sup>+</sup> area colocalized with S100 $\beta$  (not significant by Student's  $t$ -test). **e**, Electron microscopy was performed on hippocampal CA3 sections from day 7 after mock (left panel) or WNV-NS5-E218A (right panels) infection, with immune-DAB enhancement of IBA1.

Note the presence of many phagosomes and cytoplasmic inclusions within the WNV-E218A-infected microglia. Electron micrographs shown are representative of  $n = 3$  mice per group. Scale bars, 1  $\mu\text{m}$ . **f**, Immunostaining for the presynaptic marker, VGlut1, and endogenous murine IgG (mIgG) at 7 days after mock ( $n = 4$ ) or WNV-NS5-E218A ( $n = 4$ ) infection. Quantification was performed on the total per cent of mIgG staining area as well as the per cent of VGlut1<sup>+</sup> staining area colocalized with mIgG. **g**, Immunostaining for the postsynaptic marker, Homer1, and endogenous mIgG at 25 days after mock ( $n = 4$ ) or WNV-NS5-E218A-infection, which were divided into WNV-infected mice which made fewer than 8 errors on day 2 of the Barnes maze (WNV good learners,  $n = 5$ ) and WNV-infected mice which made greater than 9.5 errors on day 2 of the Barnes maze testing (WNV poor learners,  $n = 3$ ). Quantification was performed on the total per cent of mIgG staining area as well as the percent of Homer1<sup>+</sup> staining area colocalized with mIgG. Significance was determined by Student's two-tailed  $t$ -test with  $P < 0.05$  considered as significant. NS, not significant. **h**, Immunostaining and quantification of number of VGlut1 hippocampal CA3 presynaptic terminals at 7 dpi in wild-type and  $\mu\text{MT}^{-/-}$  mice. ( $*P < 0.05$ , NS, not significant, by Student's two-tailed  $t$ -test). Scale bars, 10  $\mu\text{m}$ .



**Extended Data Figure 4 | WNV infection of human hippocampal CA2/CA3 neurons with loss of synapses within the hippocampal CA1 and the entorhinal cortex.** **a**, Immunostaining of human WNV encephalitis and control post-mortem hippocampal tissue for WNV-antigen. Shown at high magnification are neuron cell bodies (arrows) and neurites (arrowheads) within the hippocampal CA2/CA3 region. **b**, **c**, Immunostaining within the hippocampal CA1 (**b**) or entorhinal

cortex (**c**) for the presynaptic marker, synaptophysin, within human WNV encephalitis and control autopsy cases. Quantification of the per cent of synaptophysin<sup>+</sup> area (hippocampal CA1  $P=0.3$ , entorhinal cortex  $P=0.11$  by two-tailed Student's  $t$ -test (not significant)). Scale bar, 20  $\mu\text{m}$ . In one WNV encephalitis patient sample, the entorhinal cortex could not be quantified because it was missing from the section.

# The bacteriophage $\phi$ 29 tail possesses a pore-forming loop for cell membrane penetration

Jingwei Xu<sup>1</sup>, Miao Gui<sup>1</sup>, Dianhong Wang<sup>1</sup> & Ye Xiang<sup>1</sup>

Most bacteriophages are tailed bacteriophages with an isometric or a prolate head attached to a long contractile, long non-contractile, or short non-contractile tail<sup>1</sup>. The tail is a complex machine that plays a central role in host cell recognition and attachment, cell wall and membrane penetration, and viral genome ejection. The mechanisms involved in the penetration of the inner host cell membrane by bacteriophage tails are not well understood. Here we describe structural and functional studies of the bacteriophage  $\phi$ 29 tail knob protein gene product 9 (gp9). The 2.0 Å crystal structure of gp9 shows that six gp9 molecules form a hexameric tube structure with six flexible hydrophobic loops blocking one end of the tube before DNA ejection. Sequence and structural analyses suggest that the loops in the tube could be membrane active. Further biochemical assays and electron microscopy structural analyses show that the six hydrophobic loops in the tube exit upon DNA ejection and form a channel that spans the lipid bilayer of the membrane and allows the release of the bacteriophage genomic DNA, suggesting that cell membrane penetration involves a pore-forming mechanism similar to that of certain non-enveloped eukaryotic viruses<sup>2–4</sup>. A search of other phage tail proteins identified similar hydrophobic loops, which indicates that a common mechanism might be used for membrane penetration by prokaryotic viruses. These findings suggest that although prokaryotic and eukaryotic viruses use apparently very different mechanisms for infection, they have evolved similar mechanisms for breaching the cell membrane.

Bacteriophage  $\phi$ 29 infects Gram-positive *Bacillus subtilis* cells via a short non-contractile tail. The  $\phi$ 29 double-stranded (ds)DNA genome is encapsulated in a prolate capsid head<sup>5–7</sup>. Both of the 5' ends of the DNA genome are covalently connected to the gene product 3 (gp3) protein<sup>6,8</sup>. The  $\phi$ 29 tail is characterized by 12 tail spikes or appendages that hang around the approximately 380 Å-long tail tube<sup>9,10</sup>. The tail tube has a gp11 protein assembly at its proximal end and a gp9 and gp13 protein assembly at its distal end<sup>11</sup>. The gp11 protein assembly, also known as the 'lower collar', has a thin tube and a bulge that is attached to the head. The lower collar tube is filled with the terminal protein gp3 and part of the genomic DNA<sup>9</sup>. The gp9 and gp13 protein assembly, also known as the tail 'knob', is a cylindrical tube whose distal end is blocked when DNA ejection has not been triggered. The tail spikes function to recognize and digest host cell wall teichoic acids and anchor the phage particle onto the host cell surface<sup>12</sup>. Gp13 is a dual-function enzyme that specifically degrades the cell wall peptidoglycan and is probably the tail protein that helps the phage tail to penetrate the thick peptidoglycan layer<sup>11,13,14</sup>. The  $\phi$ 29 tail must also penetrate the cell membrane in order to deliver the gp3-genomic dsDNA complex into the host cell cytoplasm. The mechanisms involved in membrane penetration by the  $\phi$ 29 tail and many other phage tails are unknown.

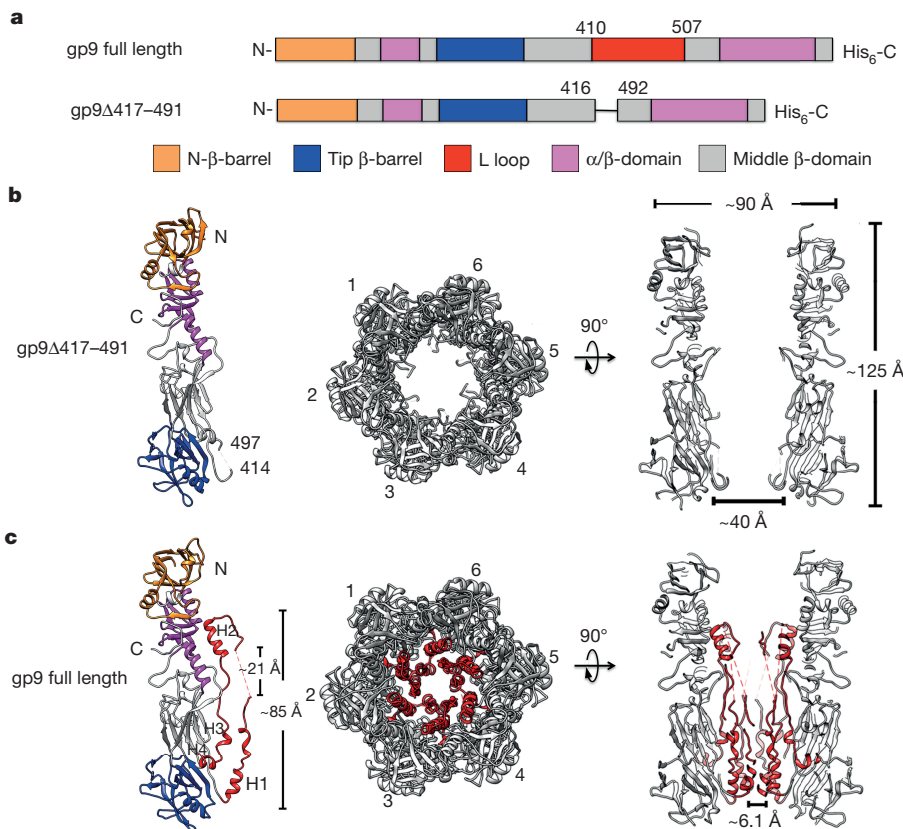
The gp9 assembly, which is located at the tip of the tail and constitutes most of the tail knob, might be associated with host cell membrane penetration by the tail. We determined the crystal structures of full-length gp9 and a mutant construct, gp9 $\Delta$ 417–491, in which a

disordered region (residues 417–491) was deleted (Fig. 1a, Methods, Extended Data Fig. 1 and Extended Data Tables 1, 2). The gp9 $\Delta$ 417–491 structure is a cylindrical tube-like homo-hexamer. The longest dimension of the tube is approximately 125 Å. The tube has an outer diameter of approximately 90 Å and an inner diameter of approximately 40 Å. The wall of the tube is approximately 25 Å thick and comprises largely  $\beta$ -sheets (Fig. 1b, c). The polypeptide chain starts from one end of the tube and ends around the middle of the tube. The N-terminal roughly 110 residues form a small  $\beta$ -barrel structure (N- $\beta$ -barrel domain) that is frequently observed in other phage tail proteins or tube-forming proteins, functioning as an adaptor to mediate the interactions between tail- and tube-forming proteins<sup>15</sup>. The polypeptide chain after the N- $\beta$ -barrel domain extends to the other end of the tube and folds into another  $\beta$ -domain (tip  $\beta$ -domain). A three-strand  $\beta$ -sheet of the tip  $\beta$ -domain bulges from the outer surface of the tube and creates features similar to the distal end of the tail knob. The remaining C-terminal region of the polypeptide chain constitutes the central part of the tube, which contains an  $\alpha/\beta$ -domain adjacent to the N- $\beta$ -barrel domain and a  $\beta$ -domain (middle  $\beta$ -domain) in which the  $\beta$ -strands are almost parallel with the tube axis (Fig. 1b, c). The structure of full-length gp9 contains a similar cylindrical hexamer, in which the deleted long loop (residues 417–491, L loop) protrudes from the inner wall near the tip  $\beta$ -domain and fills approximately two-thirds of the tube (Fig. 1c). The L loop contains several short helices and is rich in hydrophobic residues. The inner wall surface of the tube is negatively charged (Extended Data Fig. 2a). However, the surface of the L loop inside the tube is largely hydrophobic (Extended Data Fig. 2b).

Fitting of the gp9 hexameric structure into a cryo-electron microscopy (cryo-EM) map of the tail (EMDB entry 1420) unambiguously places the tip  $\beta$ -domains at the distal end of the tail and the N- $\beta$ -barrel domains in contact with the lower collar protein assembly (a correlation coefficient score of 0.673 versus 0.637 for an upside-down fitting using the program Situs<sup>16</sup>; Extended Data Fig. 3). The fitting result suggests that, like the  $\beta$ -barrel domain in other phage tails, the N- $\beta$ -barrel domain of gp9 may also function as an adaptor in mediating interactions with the lower collar protein assembly. The fitted gp9 assembly accounts for approximately three-quarters of the tail knob density. The top one-quarter, uninterpreted density of the tail knob is probably a  $\beta$ -barrel structure as well and should be part of the lower collar protein assembly. Consistent with the features of the electron microscopy structure, the fitted gp9 assembly structure has its distal end blocked by the interior L loops, leaving the proximal end empty.

The L loop in the tail tube connects two anti-parallel  $\beta$ -strands of the middle  $\beta$ -domain and has four short helices (H1–H4) and a flexible region (residues 437–457) that is highly hydrophobic and is not visible in the final structure (Fig. 1c). The polypeptide chain of the L loop extends along the tube axis for approximately 85 Å and then turns 180° near the top of the  $\alpha/\beta$ -domain, leaning against the inner tube wall. The chain falls back for approximately 70 Å and finally joins the middle  $\beta$ -domain immediately above the tip  $\beta$ -domain (Fig. 1).

<sup>1</sup>Centre for Infectious Diseases Research, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Beijing Advanced Innovation Center for Structural Biology, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China.



**Figure 1 | Overall structure of gp9.** **a**, Diagram showing the polypeptide chains of full-length gp9 and gp9 $\Delta$ 417–491. The structural domains of gp9 are shown in orange (N- $\beta$ -barrel), blue (tip  $\beta$ -barrel), grey (middle  $\beta$ -domain), purple ( $\alpha/\beta$ -domain) and red (L loop). **b**, **c**, Left, ribbon diagram of structures of gp9 $\Delta$ 417–491 (**b**) and full-length gp9 (**c**) monomers. The structural domains are coloured as in **a**. The four short

helices in the L loop structure are labelled H1–4. Middle, ribbon diagram of hexamer structures of gp9Δ417–491 (**b**) and full-length gp9 (**c**) with the L loops coloured red. The positions of the six monomers are indicated by numbers. Right, ribbon diagram of a central section of the gp9Δ417–491 (**b**) and full-length gp9 (**c**) hexamer structures with the L loops coloured red. The disordered regions of the L loops are represented by dashed lines.

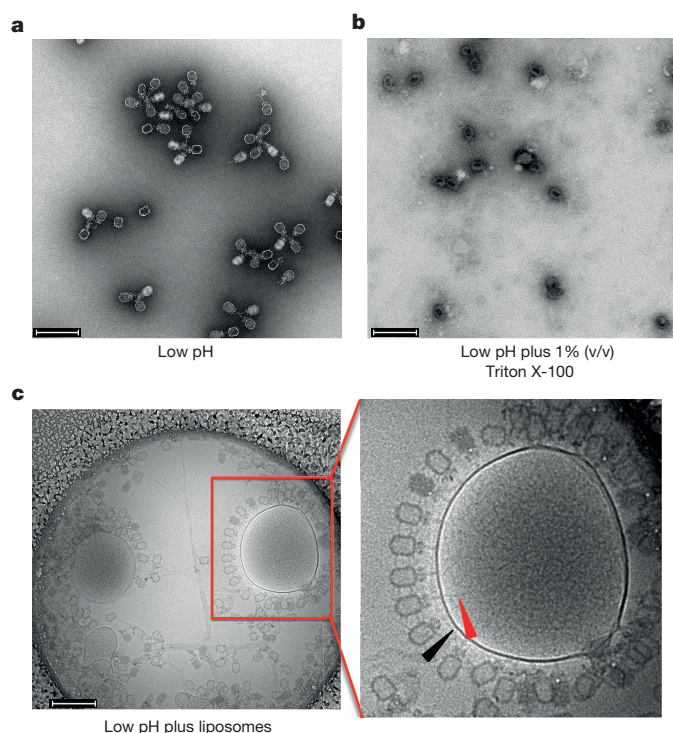
The six H1 helices near the distal end of the tube form a helix barrel around the tube axis. The missing flexible region is approximately 21 Å long and is located near the tube axis close to where the loop turns. The L loops fill most of the inner space of the tube, leaving a channel that is about 6.1 Å wide near the tube axis. The narrow channel is obviously too small for the transportation of dsDNA, which has a gyration radius of approximately 20 Å. However, as shown in the gp9Δ417–491 structure, the tube without the L loop has an inner channel that is sufficiently large (about 40 Å in diameter) for dsDNA ejection, which suggests that the L loop must exit from the tube before or simultaneously with DNA ejection. A search<sup>17</sup> for homologous sequences of the L loop shows that the flexible region of the L loop presents hydrophobic features and a sequence pattern (Fig. 2) that are similar to those of the type I fusion peptides used by certain enveloped viruses, such as HIV and influenza viruses. Hydrophobic membrane active peptides that are essential for host cell entry have also been found in non-enveloped viruses<sup>18</sup>. The search results suggest that the L loop may be membrane active and play a role in membrane penetration.

We found that the DNA ejection of bacteriophage  $\phi 29$  could be induced by low-pH buffers at around pH 4.0. Under low-pH conditions, the DNA-emptied particles tend to aggregate through the tail tips (Fig. 3a). However, the particles are dispersed in the presence of 1% (v/v) Triton X-100 (Fig. 3b), which suggests that the newly exposed surface after DNA ejection is hydrophobic. Cryo-EM studies of the low-pH-treated, Triton X-100-stabilized and DNA-emptied particles show an additional cone-shaped structure at the distal tip of the tail knob. The cone shape density extends the tail for approximately 40 Å and is connected to the tail knob through a narrow neck (Fig. 4a, b). A similar cone-shaped structure was observed in previous

cryo-EM studies of sodium-perchlorate-treated DNA-emptied phage particles<sup>9</sup>; this structure had an inner channel for dsDNA release and was not clearly assigned to any of the tail proteins. The post-DNA-ejection structures of the tail knob and the biochemical data suggest that the hydrophobic L loop exits to form the cone-shaped structure when DNA ejection is triggered. To test further the possible membrane penetration function of the L loop, we premixed mature phage particles with liposomes containing phospholipids similar to those in the *B. subtilis* cell membrane<sup>19</sup>. The phage particles were full and dispersed in the neutral pH liposome solution (Extended Data Fig. 4a). However, when DNA ejection was triggered using low-pH buffers, cryo-EM examination of the phage-liposome mixture showed that the DNA-emptied particles were able to stand on the liposome surface with their tail tips in contact with the membrane (Fig. 3c), which is similar to phage particles on an infected host cell surface<sup>20</sup>. Dense DNA densities are observed within the liposomes, suggesting that a large amount of the genomic DNA has been



**Figure 2 | Sequence alignment of the gp9 L loop and the HIV fusion peptide.** Secondary structures are shown below the alignments. Completely conserved residues are shown in white on a red background. Conserved residues are boxed. The consensus sequence is represented by H for hydrophobic, G for glycine or serine and X for any residue. The core region of the HIV fusion peptide is indicated by a red dashed box.

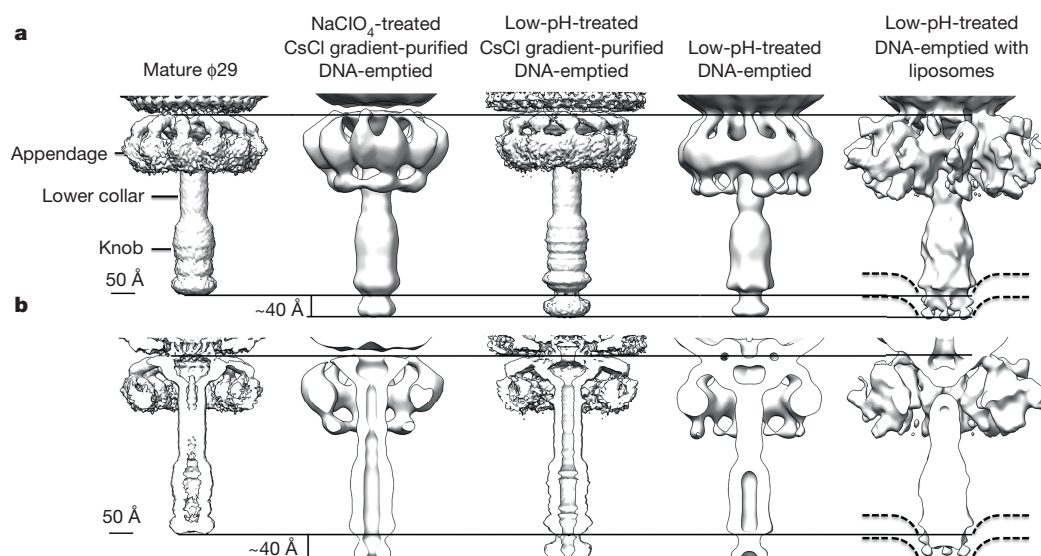


**Figure 3 | EM images of post DNA-ejection  $\phi 29$  particles.** **a**, Negatively stained image of low-pH-treated  $\phi 29$  particles showing aggregation of most particles via the tail tips. **b**, Negatively stained image of low-pH-treated, Triton-X-100-stabilized  $\phi 29$  particles showing dispersion of the particles. **c**, A cryo-EM image showing low-pH-treated  $\phi 29$  particles standing on liposomes. Dense dsDNA strands can be observed inside the liposomes. Left,  $\phi 29$  particle and liposome complexes in a grid hole ( $\sim 1.2 \mu\text{m}$  in diameter); right, enlarged view of a liposome- $\phi 29$  particle complex (boxed area in left-hand image). Black arrowhead, boundary of the liposome membrane; red arrowhead, dsDNA strand within the liposome. Scale bars, 200 nm.

ejected into the liposomes. Single-particle reconstructions of the particles on the liposomes show a similar cone-shaped structure at the distal end of the tail. The cone-shaped structure is embedded in

the membrane and spans the lipid bilayer (Fig. 4). Notably, the membrane is dented around the position where the cone-shaped structure is embedded. Compared with the sodium-perchlorate-treated particles, a small portion of the genomic dsDNA remains in the heads of the low-pH-treated particles. The low-pH-treated cone-shaped structures lack an obvious inner channel, probably because they are filled with the dsDNA genome. Further purification of the low-pH-treated particles with an isopycnic CsCl gradient completely removed the genomic dsDNA from the emptied particles. A cryo-EM reconstruction of the low-pH-treated and CsCl-gradient-purified particles was calculated and showed a cone-shaped structure with an inner channel at the distal end of the tail (Fig. 4). The pore-forming ability of the L loop was confirmed using a dye release assay (Methods and Extended Data Fig. 5). Structural modelling of the L loop conformation in the cone-shaped structure using RosettaCM<sup>21</sup> shows that the flexible region and H2 form the wall of the cone-shaped structure (Extended Data Fig. 6). Consistent with the experimental data, the modelled structure also suggests that the outer surface of the cone-shaped structure is highly hydrophobic. The height of the cone-shaped structure is approximately the same as the thickness of a lipid bilayer. Notably, we found that the gp9 L loop could also penetrate a liposome membrane containing phospholipids similar to those in the eukaryotic cell membrane (Extended Data Fig. 4b). Blast searches for similar loops in phage tail proteins suggested that this mechanism might be used for cell membrane penetration by many phages (Extended Data Fig. 7), including bacteriophage T4, which has a long contractile tail.

Viruses must breach the physical membrane barrier to deliver their genome into the host cell cytoplasm. Most enveloped eukaryotic viruses insert a hydrophobic fusion peptide into the host membrane by using conformational changes in the fusion-peptide-related structure to mediate the fusion of the cell membrane and the viral membrane, thus overcoming the host membrane barrier. Non-enveloped eukaryotic viruses use a membrane-active peptide to penetrate membranes through a pore-forming or a local cellular membrane disruption mechanism<sup>18</sup>. Our identification of a hydrophobic, pore-forming, membrane-active peptide in a prokaryotic virus suggests that prokaryotic and eukaryotic viruses share common mechanisms for membrane penetration, possibly as a result of convergent evolution.



**Figure 4 | Pre- and post-DNA-ejection structures of the  $\phi 29$  tail.** Comparison of the cryo-EM densities of DNA-filled mature,  $\text{NaClO}_4$ -treated and CsCl-gradient-purified, low-pH-treated and CsCl-gradient-purified, low-pH-treated and low-pH-treated with liposomes tails. The DNA-emptied particles have an additional cone-shaped density at their

distal ends. **a**, Surface-shaded view. **b**, Cross-section of the cryo-EM densities. The lipid bilayer has weaker densities than that of the phage tail and is not visible at a contouring level of  $3.0\sigma$ . The boundary of the lipid bilayer is indicated with dashed lines.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 12 October 2015; accepted 14 April 2016.**

**Published online 15 June 2016.**

- Ackermann, H. W. Bacteriophage observations and evolution. *Res. Microbiol.* **154**, 245–251 (2003).
- Panjwani, A. *et al.* Capsid protein VP4 of human rhinovirus induces membrane permeability by the formation of a size-selective multimeric pore. *PLoS Pathog.* **10**, e1004294 (2014).
- Shukla, A., Padhi, A. K., Gomes, J. & Banerjee, M. The VP4 peptide of hepatitis A virus ruptures membranes through formation of discrete pores. *J. Virol.* **88**, 12409–12421 (2014).
- Galloux, M. *et al.* NMR structure of a viral peptide inserted in artificial membranes: a view on the early steps of the birnavirus entry process. *J. Biol. Chem.* **285**, 19409–19421 (2010).
- Tao, Y. *et al.* Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95**, 431–437 (1998).
- Anderson, D. L. & Reilly, B. E. in *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics* (eds Sonenshein, A. L., Hoch, J. A. & Losick, R.) 859–867 (American Society for Microbiology, 1993).
- Morais, M. C. *et al.* Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of  $\phi 29$ . *Mol. Cell* **18**, 149–159 (2005).
- Meijer, W. J., Horcajadas, J. A. & Salas, M.  $\phi 29$  family of phages. *Microbiol. Mol. Biol. Rev.* **65**, 261–287 (2001).
- Xiang, Y. *et al.* Structural changes of bacteriophage  $\phi 29$  upon DNA packaging and release. *EMBO J.* **25**, 5229–5239 (2006).
- Tang, J. *et al.* DNA poised for release in bacteriophage  $\phi 29$ . *Structure* **16**, 935–943 (2008).
- Xiang, Y. *et al.* Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage  $\phi 29$  tail. *Proc. Natl Acad. Sci. USA* **105**, 9552–9557 (2008).
- Xiang, Y. *et al.* Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. *Mol. Cell* **34**, 375–386 (2009).
- Cohen, D. N. *et al.* Shared catalysis in virus entry and bacterial cell wall depolymerization. *J. Mol. Biol.* **387**, 607–618 (2009).
- Cohen, D. N., Erickson, S. E., Xiang, Y., Rossmann, M. G. & Anderson, D. L. Multifunctional roles of a bacteriophage  $\phi 29$  morphogenetic factor in assembly and infection. *J. Mol. Biol.* **378**, 804–817 (2008).
- Cardarelli, L. *et al.* Phages have adapted the same protein fold to fulfill multiple functions in virion assembly. *Proc. Natl Acad. Sci. USA* **107**, 14384–14389 (2010).
- Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys. Rev.* **2**, 21–27 (2010).
- Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2011).
- Johnson, J. E. & Vogt, P. K. Cell entry by non-enveloped viruses. *Curr. Top. Microbiol. Immunol.* **343**, v–vii (2010).
- Doan, T. *et al.* FisB mediates membrane fission during sporulation in *Bacillus subtilis*. *Genes Dev.* **27**, 322–334 (2013).
- Xiang, Y. & Rossmann, M. G. Structure of bacteriophage  $\phi 29$  head fibers has a supercoiled triple repeating helix-turn-helix motif. *Proc. Natl Acad. Sci. USA* **108**, 4806–4810 (2011).
- Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).

**Acknowledgements** We thank L. Q. Zhang, N. Yan, H. T. Li, S. L. Fan, N. Gao, C. Z. Zhou, D. L. Anderson and M. G. Rossmann for support; the Tsinghua University Branch of the China National Center for Protein Sciences for the facility support; and the staff at the Shanghai Synchrotron Research Facility beam line BL17U for assistance with data collection. This work was supported by funds from the 973 program (2015CB910102), the National Natural Science Foundation of China (31470721 and 81550001), the Junior Thousand Talents Program of China (20131770418) and the Beijing Advanced Innovation Center for Structural Biology to Y.X.

**Author Contributions** J.X. and Y.X. designed the research; J.X., M.G., D.W. and Y.X. performed the experiments; J.X. and Y.X. analysed the data and wrote the paper; and all authors contributed to the editing of the manuscript.

**Author Information** The atomic coordinates and structure factor files have been deposited into the Protein Data Bank (PDB) under accession numbers 5FB4, 5FB5 and 5FEI. The electron microscopy maps have been deposited into the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-6556, EMD-6557 and EMD-6558. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.X. ([yxiang@mail.tsinghua.edu.cn](mailto:yxiang@mail.tsinghua.edu.cn)).

## METHODS

**Protein expression and purification.** Full-length gp9 consists of 599 residues. It took approximately half a year for the full-length protein crystals to grow. An SDS-PAGE gel analysis of the protein in the crystallization wells showed that most of the full-length protein was degraded to form several smaller fragments. A mass spectrometry analysis of the fragments indicated two cleavage sites at residues 398 and 501. Additionally, secondary structure and disordered region predictions showed that residues 414–499 did not form any secondary structure and were disordered (Extended Data Fig. 1). A series of constructs were produced to delete the disordered region. Among the tested constructs, construct gp9 $\Delta$ 417–491 (Fig. 1a), which had the best expression level in *Escherichia coli* cells, was selected for large-scale recombinant protein production, purification and crystallization. Gene 9 was PCR-amplified from the genomic DNA of bacteriophage  $\phi$ 29 using the following primers: 5'-CCCATGGCATATGTACCATATCAGGAACG-3' and 5'-CCGCTC GAGTCAGTGGTGGTGGTGGTGGTGCCTCAATTCATTCTCGACGC-3'. The genomic DNA of bacteriophage  $\phi$ 29 was prepared by heating the mature  $\phi$ 29 phage ( $5 \times 10^{12}$  particles per ml) at 96°C for 10 min. One microlitre of the heat-treated phage solution was then used as a template for each PCR reaction. The purified PCR products were digested using the endonucleases NcoI and XhoI and then cloned into pET28b (Novagen) using the NcoI–XhoI sites and placing a His<sub>6</sub> tag on the C terminus of the recombinant protein. The recombinant p gene was PCR-amplified from the pET28b-gp9 plasmid using the following primers: 5'-AAGGAGGAAGCAGGTATGGCATATGTACCATATCAGGAACG-3' (forward) and 5'-CTGTGCGTGCTCCATCAGTGGTGGTGGTGGTGGTGCCTCAATTC ATTCTCGACGC-3' (reverse). The PCR products were cloned into pDG (*Bacillus* centre: <http://bgsc.org>) using a ligation-free method<sup>22</sup>. The gp9 loop deletion mutants were generated using two-step overlapping PCR with the primers listed in Extended Data Table 1. The PCR products were digested with NcoI and XhoI and then cloned into pET28b.

The pDG-gp9 plasmids were transformed into *Bacillus subtilis* cells using the natural competence of the cells that developed at the end of the logarithmic growth phase under semi-starvation conditions (see [http://wiki.biol.uw.edu.pl/t/img\\_auth.php/b/b5/Bacillus\\_subtilis\\_competent\\_cells.pdf](http://wiki.biol.uw.edu.pl/t/img_auth.php/b/b5/Bacillus_subtilis_competent_cells.pdf) for the protocol details)<sup>23</sup>. Colonies containing the pDG-gp9 plasmids were selected and cultivated at 37°C until they reached an OD<sub>600</sub> value of  $\sim 1$ . Then, recombinant full-length gp9 was induced for expression with 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). The recombinant gp9 mutants were expressed in *E. coli* BL21(DE3) cells at 16°C following the standard protocol for IPTG-induced protein expression in *E. coli* cells (see the Novagen pET system handbook). The produced recombinant protein was affinity purified using cobalt-charged BD TALON resins and eluted from the cobalt beads using an elution buffer containing 100 mM sodium phosphate at pH 8.0, 300 mM sodium chloride and 200 mM imidazole. The eluted protein was concentrated and further purified with a Superdex 200 column (GE) running in a buffer containing 20 mM Tris pH 8.0, 300 mM sodium chloride and 1 mM DTT. The peak fractions were collected and concentrated to  $\sim 10$  mg ml<sup>-1</sup> for crystal screening. Size-exclusion chromatography purification of full-length gp9 indicated that this protein exists mainly as a monomer, whereas the gp9 $\Delta$ 417–491 mutant exists mainly as a multimer, probably a hexamer, in solution.

**Crystallization.** All crystals were obtained by hanging-drop vapour diffusion at 20°C using 2  $\mu$ l protein (10 mg ml<sup>-1</sup>) mixed with an equal volume of well solution. Crystals of full-length gp9 were grown in 0.1 M sodium acetate pH 4.6 and 2.0 M ammonium sulphate. It took approximately half a year for the crystals to grow. The crystals were soaked for 30 s in the well solution containing a final concentration of 20% (v/v) glycerol to flash freeze in liquid N<sub>2</sub>. The gp9 $\Delta$ 417–491 crystals were grown under two different conditions. Plate-shaped crystals were grown in 0.94% (v/v) ethanol and 10% PEG-400 in a HEPES buffer pH 7.5 and 2 M magnesium chloride. Diamond-shaped crystals were grown in 1.2 M lithium sulphate and 10% (v/v) PEG-400 in a sodium acetate buffer pH 4.4. The gp9 $\Delta$ 417–491 crystals were soaked for 30 s in the well solution containing 10% (v/v) glycerol to flash freeze in liquid N<sub>2</sub>. The mercury derivative of gp9 $\Delta$ 417–491 was prepared by soaking the diamond-shaped crystals in the cryo-well solution containing 10 mM K<sub>2</sub>HgI<sub>4</sub> for 2 h.

**Phage production, cryo-EM data collection and image processing.** Fibred  $\phi$ 29 particles were produced in *Bacillus subtilis* su44<sup>+</sup> cells infected with the mutants sus16(300)-sus14(1241) and purified by centrifugation in an isopycnic 65% (w/v) CsCl gradient. The purified phage particles ( $5 \times 10^{12}$  particles per ml) were in a buffer containing 50 mM Tris-HCl pH 7.8, 100 mM NaCl and 10 mM MgCl<sub>2</sub>. Liposomes containing phospholipids similar to those in the *B. subtilis* cell membrane were prepared from a chloroform solution consisting of phosphatidyl-ethanolamine (PE), phosphatidyl-DL-glycerol (PG), phosphatidyl-choline (PC) and cholesterol (CL) at a molar ratio of 5:6:2:9. Liposomes containing phospholipids similar to those in the eukaryotic cell membrane were prepared from a chloroform solution consisting of PC, PE, CL and sphingomyelin (SPH) at a molar ratio of

1:1:3:1. The phospholipid chloroform solution was slowly blow-dried using N<sub>2</sub> to form films consisting of stacked phospholipid bilayers. The films were further dried in a vacuum desiccator for 24 h and then re-suspended in a buffer containing 25 mM HEPES pH 7.4, 100 mM KCl, 10% glycerol, and 1 mM DTT. The phospholipid bilayer emulsion was frozen and thawed several times and then sonicated for approximately 120 s until the solution cleared. The sonicated solution was passed through a 0.1- $\mu$ m filter 10 times to form liposomes with a uniform size of approximately 1,000 Å. The *in vitro* ejection of the  $\phi$ 29 genome was triggered by adjusting the buffer pH to  $\sim 4.2$  with a low-pH buffer containing 0.1 M sodium acetate pH 4.0 and 300 mM ammonium sulphate in the presence of either 0.8% (v/v) Triton X-100 (phage, low-pH buffer and 10% (v/v) Triton X-100 at a volume ratio of 1:10:1) or 5 mM liposomes (phage, low-pH buffer and liposomes at a volume ratio of 1:10:10), followed by incubation at 37°C for 20 h. The detergent-stabilized samples were gradually transferred to a buffer containing 10 mg ml<sup>-1</sup> amphipol A8-35 (Anatrace, A835) on an affinity grid and flash frozen in liquid ethane at 100 K. The affinity grids were prepared based on a recently published protocol<sup>24</sup>. The activated carbon film grids were covalently bonded to the anti- $\phi$ 29 gp9 polyclonal antibody. The low-pH-treated  $\phi$ 29-liposome complex was directly flash frozen in liquid ethane at 100 K using normal 400-mesh holey carbon grids (Quantifoil, 1.2  $\mu$ m  $\times$  1.3  $\mu$ m). To completely remove the genomic DNA, the low-pH-treated particles were further purified in an isopycnic 65% (w/v) CsCl gradient with 0.02% (v/v) Triton X-100. The CsCl gradient purified sample was directly frozen in liquid ethane at 100 K using 400-mesh thin carbon coated holey carbon grids (Lantuo Jiangsu, China, 2  $\mu$ m  $\times$  2  $\mu$ m). A total of 483 cryo-EM images of the phage in amphipol A8-35 were recorded on a K2 summit detector at a nominal magnification of 22,500 (which yields a calibrated pixel size of 1.32 Å) using an FEI Titan Krios transmission electron microscope operated at 300 kV. A total of 150 cryo-EM images of the phage-liposome complex were recorded on a CCD camera at a nominal magnification of 29,000 (which yields a calibrated pixel size of 3.02 Å) using an FEI F20 transmission electron microscope operated at 200 kV. A total of 945 images of the low-pH-treated and CsCl-gradient-purified sample were recorded on a Falcon II direct electron detector at a nominal magnification of 58,000 (which yields a calibrated pixel size of 1.97 Å) using an FEI Tecnai Arctic transmission electron microscope operated at 200 kV. Individual phage particle images were selected and boxed with the program EMAN2 (<http://blake.bcm.edu/emanwiki/EMAN2>). The contrast transfer function (CTF) parameters were determined with the 'ctffit' routine in the EMAN package<sup>25</sup>. Only the phases were corrected for the observed image data using the determined CTF parameters. The initial orientations and centres of the tails were determined by phage head reconstructions with the program EMAN assuming five-fold symmetry with the symmetry axis corresponding to the z-axis. The tail orientations were determined by searching the rotation around the z-axis and assuming six-fold symmetry for the reconstructions. The particle numbers used in the final calculation of the low-pH-treated, low-pH-treated and CsCl-gradient-purified, and low-pH-treated with liposomes samples were 3,420, 24,544 and 802, respectively. The resolutions of the final reconstructed detergent-stabilized tail, detergent-stabilized and CsCl-gradient-purified tail, and the tail-liposome complex maps were estimated to be  $\sim 15.4$  Å, 10.1 Å and 34.5 Å, respectively, using the gold standard<sup>26</sup>.

**Fluorescent dye release.** Liposomes containing calcein (Sigma, C0875) were prepared using the same phospholipids (PE, PG, PC and CL at a molar ratio of 5:6:2:9) and a procedure similar to that described for the cryo-EM sample preparation but with a re-suspension buffer containing 10 mM sodium phosphate pH 7.4, 30 mM NaCl (0.1  $\times$  PBS) and 120 mM calcein. Free calcein that was not encapsulated into the liposomes was removed using a 5-ml HiTrap desalting column (GE) running in a buffer containing 10 mM sodium phosphate pH 7.4 and 30 mM NaCl. Twenty microlitres of calcein-containing liposomes was mixed with an equal volume of the low-pH buffer containing 0.1 M sodium acetate pH 4.0 and 300 mM ammonium sulphate. After adding 2  $\mu$ l of phage solution (at a concentration of  $5 \times 10^{12}$  particles per ml) into liposomes pre-incubated at 37°C, 2  $\mu$ l of the mixture was collected every 5 min. Each aliquot was immediately diluted into 1 ml of 0.1  $\times$  PBS buffer and the fluorescence signal of the dilution was measured using a Nanodrop 3300 spectrofluorometer with the excitation and emission wavelengths set to 470 nm and 520 nm, respectively. As a control, the calcein-containing liposomes mixed with the phage and a neutral-pH buffer containing 0.1 M sodium phosphate pH 7.4 and 300 mM ammonium sulphate was measured using the same procedure. To determine the background fluorescence, the calcein-containing liposomes mixed with the low-pH buffer alone were also measured using the same procedure. All measurements were repeated at least three times. The data used for the generation of Extended Data Fig. 5 were the mean values of the three independent measurements.

**Modelling of the post-ejection conformation.** Modelling of the post-ejection conformation of the tail knob was performed using the program RosettaCM<sup>21</sup> and the MDFF package<sup>27</sup>. Only the L loops were left free to move during the modelling,

whereas the rest of the structure and the interactions between the monomers were maintained with harmonic restraints.

**X-ray data collection, processing, structure determination, refinement and analysis.** X-ray diffraction data were collected using synchrotron radiation at the Shanghai Radiation Facility beamline BL17U (Extended Data Table 2). The plate-shaped crystals of gp9 $\Delta$ 417–491 diffracted to  $\sim 2.0$  Å and belong to space group *I*222, with three molecules in the asymmetric unit and cell parameters of  $a = 94.6$  Å,  $b = 135.2$  Å and  $c = 313.4$  Å. The diamond-shaped crystals of gp9 $\Delta$ 417–491 diffracted to  $\sim 2.6$  Å and belong to space group *P*<sub>2</sub><sub>1</sub><sub>3</sub>, with two molecules in the asymmetric unit and cell parameters of  $a = b = c = 184.7$  Å. Crystals of full-length gp9 diffracted to  $\sim 3.5$  Å and belong to space group *P*<sub>2</sub><sub>1</sub><sub>3</sub>, with two molecules in the asymmetric unit and cell parameters of  $a = b = c = 183.4$  Å. The data were integrated and scaled with the HKL2000 suite (Extended Data Table 2)<sup>28</sup>.

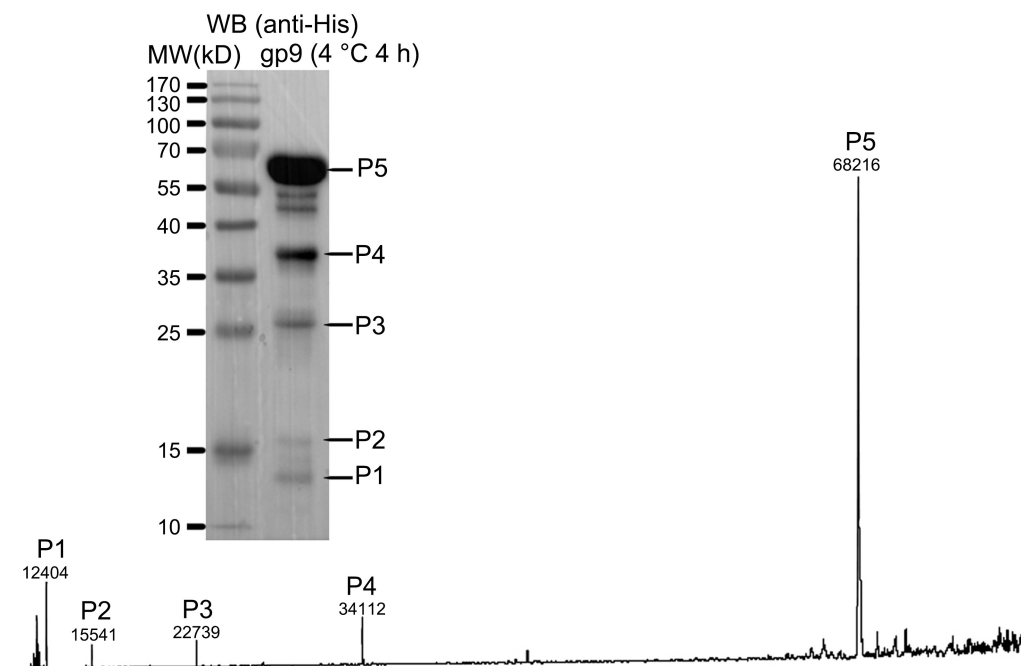
The structure of gp9 $\Delta$ 417–491 was determined by SAD using the diamond-shaped crystals and the anomalous signals of mercury atoms as measured at a wavelength of 0.97923 Å. Heavy atom sites were located using the program SHELX<sup>29</sup>. The heavy atom parameters were refined and the initial phases were calculated using reflections in the resolution region of 50 Å to 2.6 Å with the program SHARP<sup>30</sup>. The calculated phases were gradually improved and extended to a higher resolution using solvent density flattening with the program DM<sup>31</sup>. The resulting electron density maps are of good quality, in which most of the residues can be clearly recognized. The structures of full-length gp9 and gp9 $\Delta$ 417–491 in the plate-shaped crystals were determined by molecular replacement<sup>32</sup>. The full-length gp9 structure was refined to a  $R_{\text{work}}/R_{\text{free}}$  of 0.201/0.229. Of all the residues, 93% are in the most favoured regions of the Ramachandran plot and 1.1% (10 of 1140) of the residues are in the disallowed regions. The gp9 $\Delta$ 417–491 structure in the diamond-shaped crystals was refined to a  $R_{\text{work}}/R_{\text{free}}$  of 0.166/0.201. Of all the residues, 95% are in the most favoured regions of the Ramachandran plot and 0.3% (3 of 1021) of the residues are in the disallowed regions. The gp9 $\Delta$ 417–491 structure in the plate-shaped crystals was refined to a  $R_{\text{work}}/R_{\text{free}}$  of 0.153/0.183. Of all the residues, 97% are in the most favoured regions of the Ramachandran plot and no residue is in the disallowed regions. The densities of residues 437–457 of full-length gp9 are not visible in the final 3.5 Å electron density map.

The program COOT was used for model building and for making adjustments<sup>33</sup>. The program PHENIX was used to refine the structure (Extended Data Table 2)<sup>34</sup>. The program SITUS<sup>16</sup> was used to fit the structure into the six-fold-averaged cryo-EM density of the tail knob in the reconstruction of the  $\phi$ 29 mature virus<sup>10</sup>. Some of the figures were prepared with the programs Chimera<sup>35</sup> and ESPript<sup>36</sup>. Cryo-EM maps used in Fig. 4 are DNA-filled mature (low-pass filtered to a resolution of  $1/12$  Å<sup>−1</sup> and at a contouring level of  $4.0\sigma$ ), sodium-perchlorate-treated

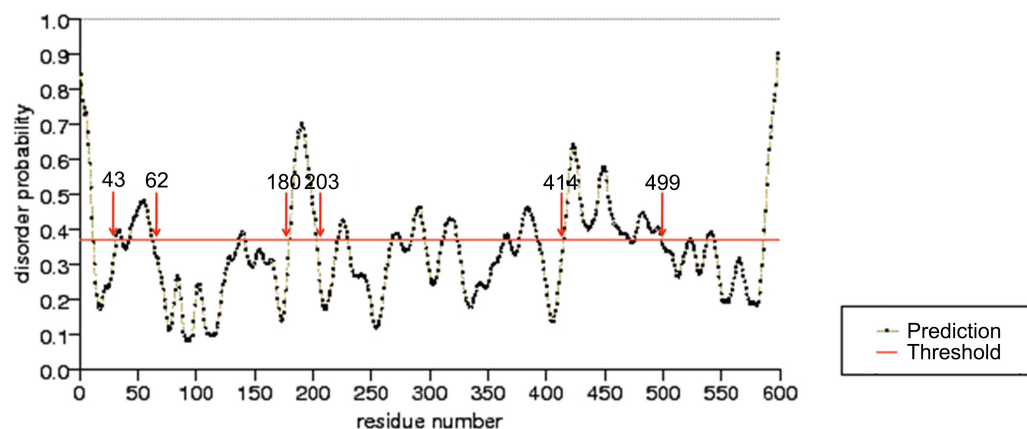
and CsCl-gradient-purified (low-pass filtered to a resolution of  $1/25$  Å<sup>−1</sup> and at a contouring level of  $2.0\sigma$ ), low-pH-treated and CsCl-gradient-purified (low-pass filtered to a resolution of  $1/10$  Å<sup>−1</sup> and at a contouring level of  $3.9\sigma$ ), low-pH-treated (low-pass filtered to a resolution of  $1/15$  Å<sup>−1</sup> and at a contouring level of  $2.1\sigma$ ) and low-pH-treated with liposomes (low-pass filtered to a resolution of  $1/26$  Å<sup>−1</sup> and at a contouring level of  $3.0\sigma$ ) tails.

22. Joseph, P., Fantino, J. R., Herbaud, M. L. & Denizot, F. Rapid orientated cloning in a shuttle vector allowing modulated gene expression in *Bacillus subtilis*. *FEMS Microbiol. Lett.* **205**, 91–97 (2001).
23. Jarmer, H., Berka, R., Knudsen, S. & Saxild, H. H. Transcriptome analysis documents induced competence of *Bacillus subtilis* during nitrogen limiting conditions. *FEMS Microbiol. Lett.* **206**, 197–200 (2002).
24. Llaguno, M. C. et al. Chemically functionalized carbon films for single molecule imaging. *J. Struct. Biol.* **185**, 405–417 (2014).
25. Ludtke, S. J. 3-D structures of macromolecules using single-particle analysis in EMAN. *Methods Mol. Biol.* **673**, 157–173 (2010).
26. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
27. Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174–180 (2009).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
29. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
30. Bricogne, G., Vonnrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
31. Cowtan, K. D., Zhang, K. Y. J. & Main, P. In *International Tables for Crystallography* Vol. F (eds Rossmann, M. G. & Arnold, E.) 507–510 (Kluwer Academic Publishers, 2000).
32. Rossmann, M. G. & Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr. D* **15**, 24–31 (1962).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
34. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
35. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
36. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
37. Ishida, T. & Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460–W464 (2007).

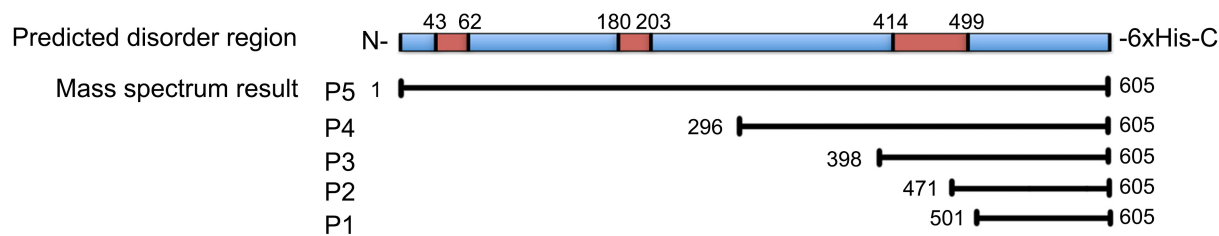
**a**



**b**

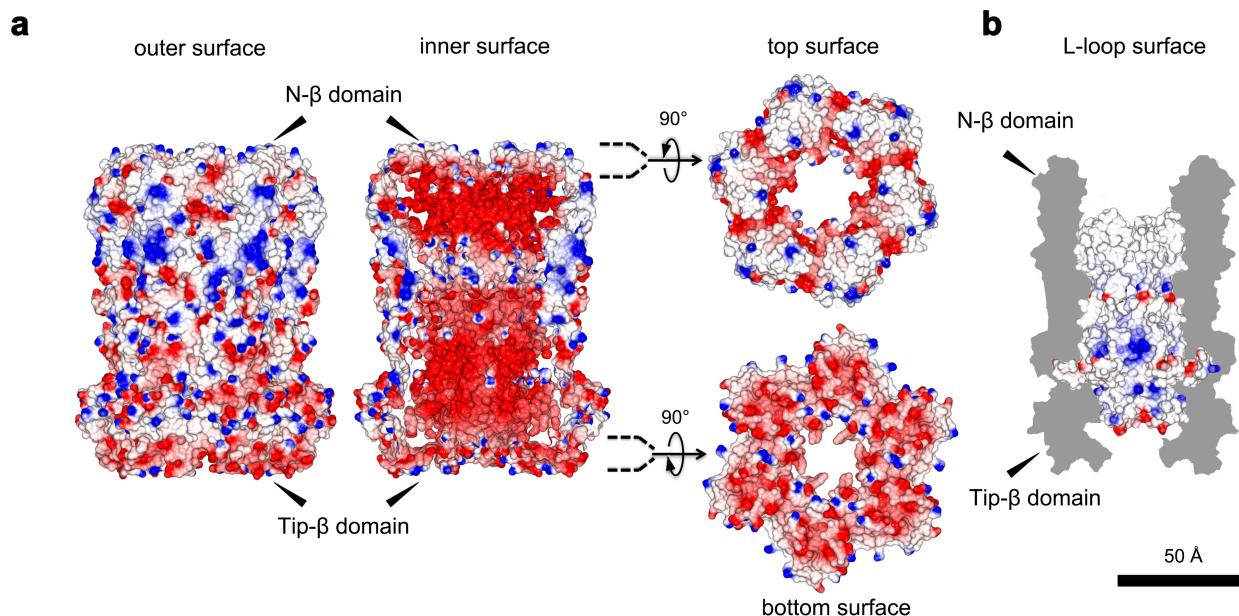


**c**



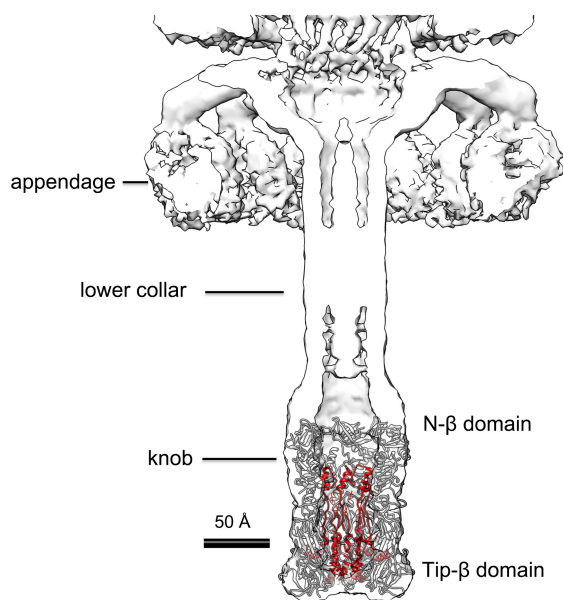
**Extended Data Figure 1 | Mass spectrometry and sequence analysis showing the disordered region of gp9.** **a**, Western blotting and mass spectrometry analysis of the recombinant full-length and degraded gp9 proteins. Positions of the full-length and degraded gp9 proteins are indicated with arrows and labelled with P5–P1 in the western blotting image. Corresponding peaks in the mass spectra are labelled in the same order as in the western blot. **b**, Disordered region analysis of gp9 using the

online tool ProDOS<sup>37</sup>. The threshold (red line) was set based on a false positive rate of 0.15. Residues with a disorder probability of more than 0.38 are considered disordered. The start and end positions of several predicted long disordered regions are indicated with red arrows. **c**, Diagrams showing the location of the predicted disordered region in the sequence and the fragments determined by the mass spectrometry analysis.

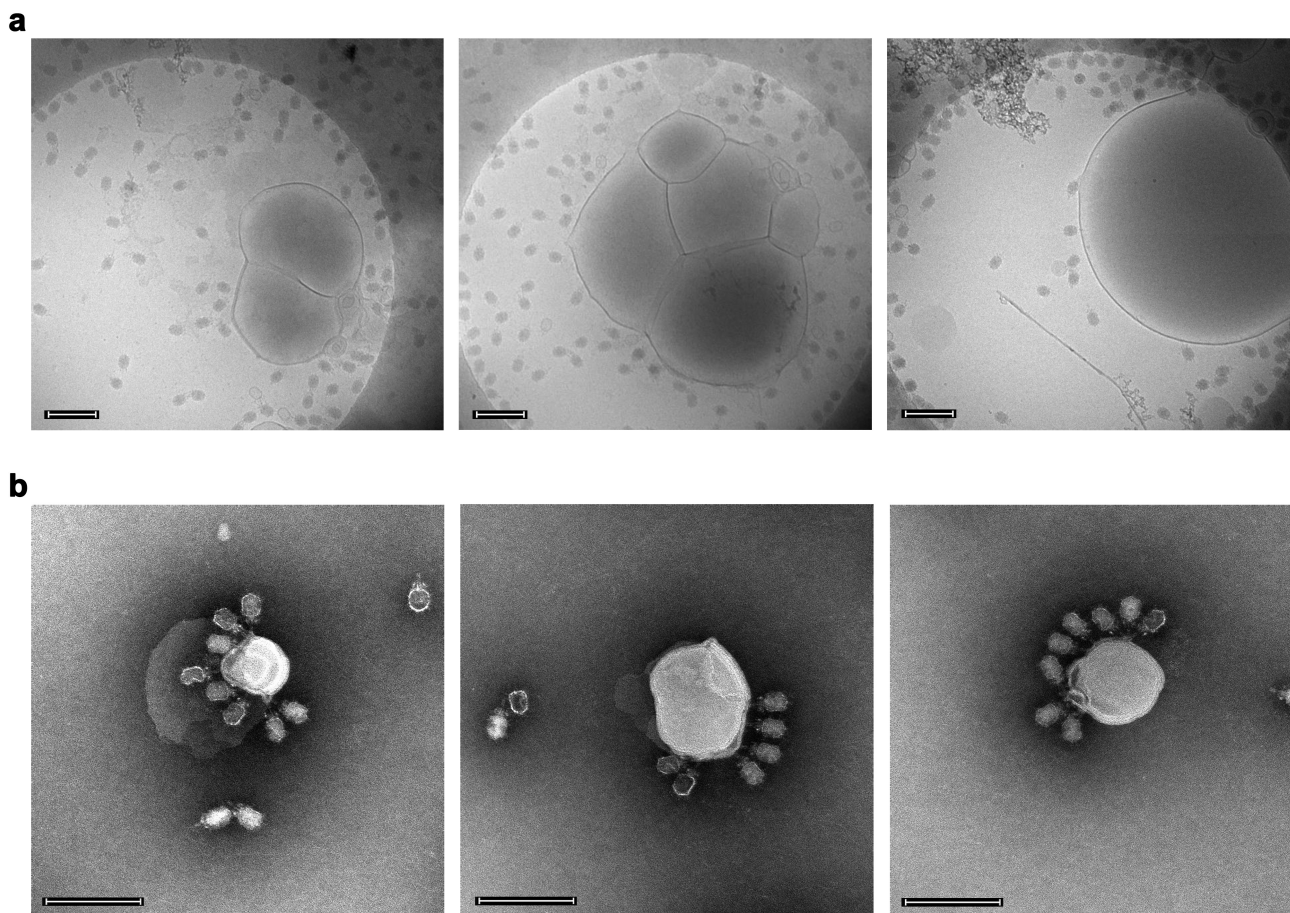


**Extended Data Figure 2 | Surface charge distribution of gp9. a**, Diagram showing the surface charge distribution of gp9 $\Delta$ 417–491. Negative and positive electrostatic potentials are coloured red and blue, respectively. The monomers within the hexamer interact with each other primarily through hydrogen bonds and hydrophobic interactions. The outer cylindrical surface of the tube is hydrophilic. The top and bottom surfaces of the tube have substantially different features. The top surface on the

N- $\beta$ -barrel domain side is largely hydrophobic, whereas the bottom surface on the tip  $\beta$ -domain side is rich in negatively charged residues. **b**, Diagram showing the surface charge distribution of the interior L loops. Negative and positive electrostatic potentials are coloured as in **a**. A thin central slice of the tube is shown and coloured dark grey to indicate the tube boundary.

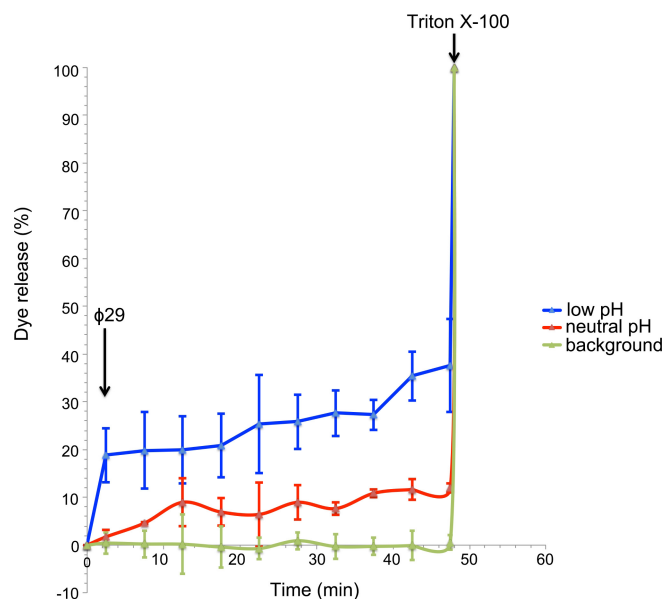


**Extended Data Figure 3 | Fitted gp9 structure in a cryo-EM map of the tail.** Diagram showing the gp9 hexamer structure fitted in a cryo-EM map of the tail contoured at  $4.0\sigma$ . An  $\sim 45$ -Å-thick cross-section is shown. The gp9 hexamer structure is shown as a ribbon representation with the L loops coloured red. The density map is shown as a solid surface representation and coloured semi-transparent grey.

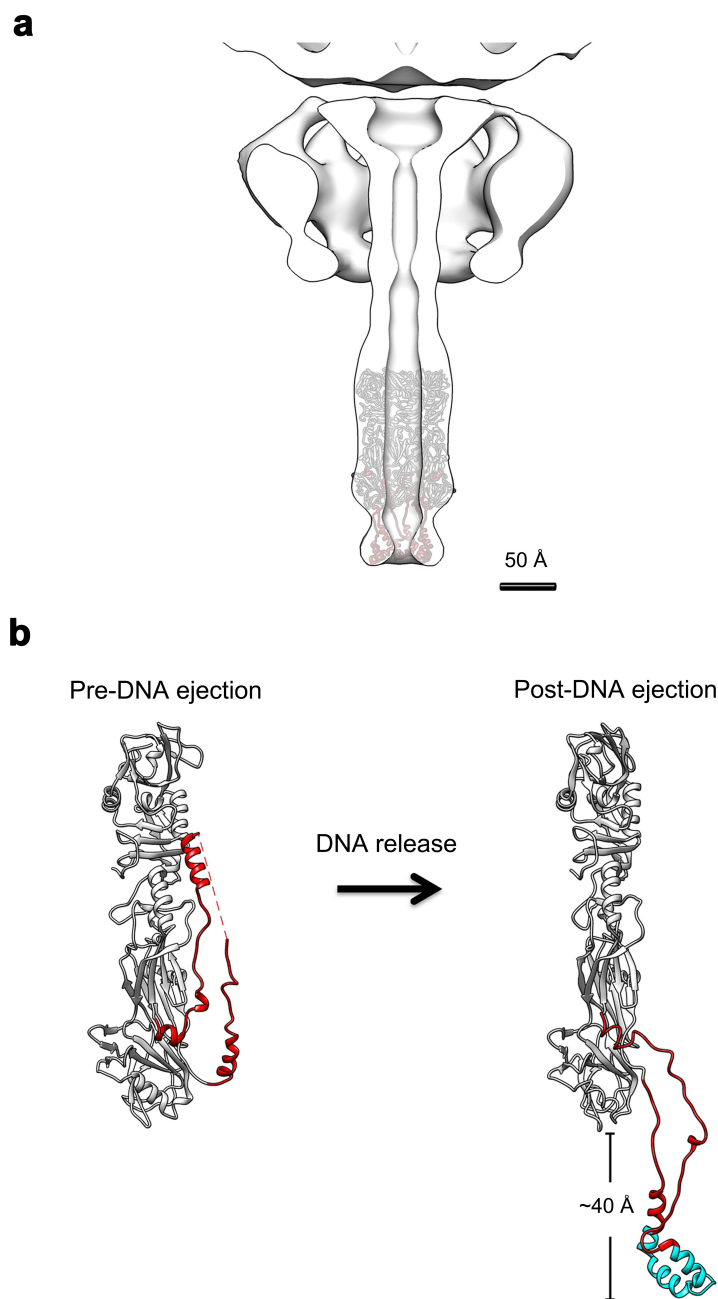


**Extended Data Figure 4 | Electron microscopy images showing the mature  $\phi 29$ -liposome complex. a,** Cryo-EM images of mature  $\phi 29$  particles with *B. subtilis* cell membrane-like liposomes at neutral pH. **b,** Negatively stained electron microscopy images showing the

low-pH-treated, DNA-emptied  $\phi 29$  particles aggregating around a liposome that contains lipids similar to those in the eukaryotic cell membrane. Scale bars, 200 nm.

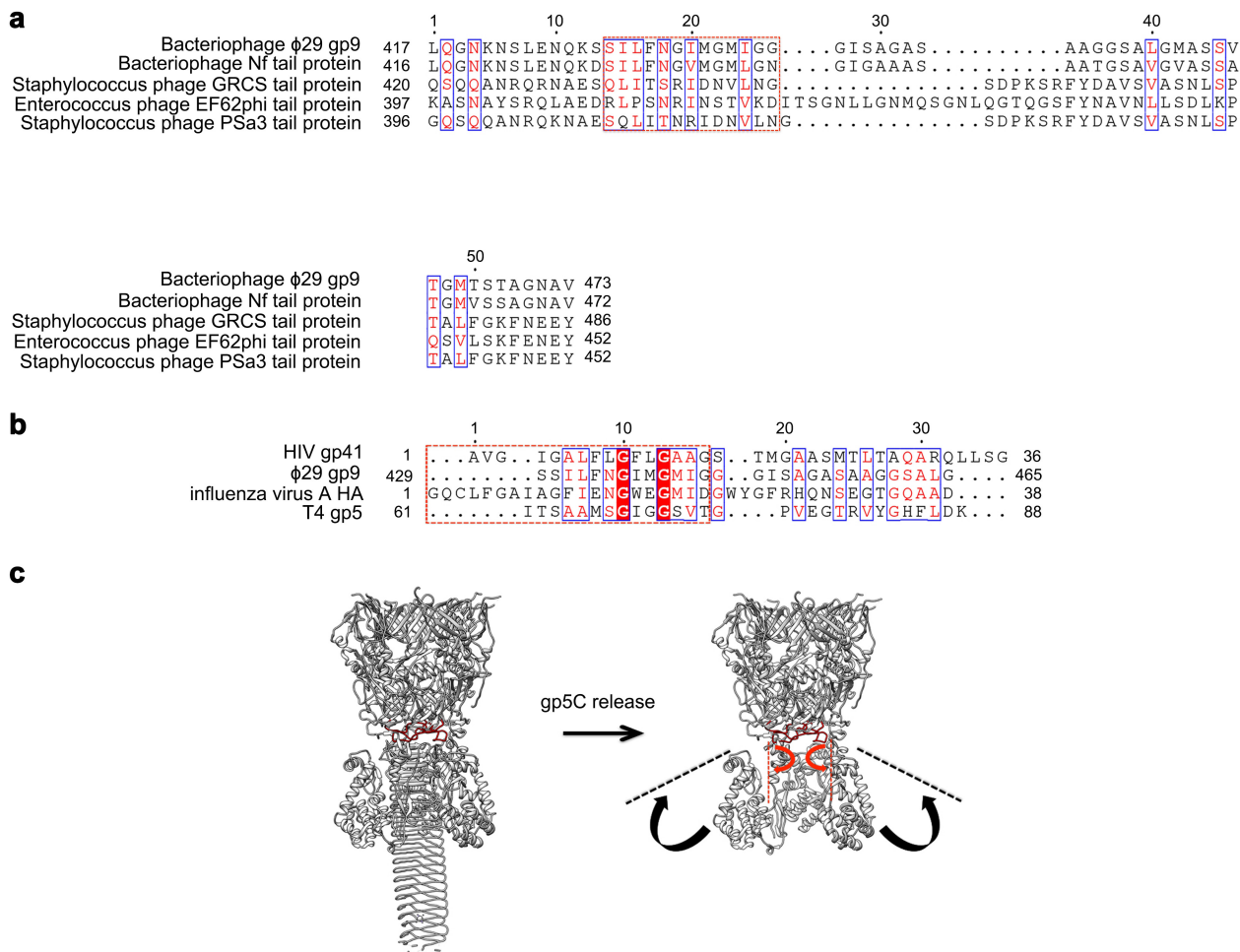


**Extended Data Figure 5 | Release of calcein from liposomes induced by  $\phi 29$ .** Time-course release of calcein from liposomes induced by  $\phi 29$  in a low- or neutral-pH buffer is shown in solid lines with blue (low pH) or red (neutral pH) triangles representing the data points. Percentages of the calcein released from the liposomes were calculated from  $(F_d - F_0) / (F_{\max} - F_0)$  where  $F_d$  is the measured fluorescence value of the low or neutral pH sample,  $F_0$  is the background fluorescence and  $F_{\max}$  is the maximum fluorescence value measured after adding Triton X-100 at a final concentration of 0.02% (v/v).  $F_0$  (the  $x$  axis) is generated from a linear least square fitting of the fluorescence values measured on a liposome-low-pH buffer mixture without adding the phage. The data used for the linear least square fitting of the background fluorescence are represented by green triangles. The data are expressed as the mean  $\pm$  s.d. of three independent measurements.



**Extended Data Figure 6 | Structure models of the pre- and post-DNA-ejection L loops.** **a**, Post-DNA-ejection L loop structure modelled with RosettaCM using the electron microscopy map as a restraint and fitted into the post-DNA-ejection tail electron microscopy map. The gp9 structure is shown in a ribbon representation. The L loops are coloured red. The rest of the structure is coloured grey. **b**, Structural comparison

of the pre- and post-DNA-ejection L loop structures. Left, pre-DNA-ejection gp9 structure shown in a ribbon representation with the L loops coloured red. Right, post-DNA-ejection gp9 structure shown in a ribbon representation with the modelled structure of the flexible region coloured cyan and the rest of the L loop coloured red.



**Extended Data Figure 7 | Sequence alignments of the gp9 L-loop-like peptides in phage tail proteins.** **a**, Sequence alignment of the gp9 L-loop and L-loop-like peptides from short non-contractile tails. Conserved residues are boxed and coloured red. The core hydrophobic region of the L loop is indicated by a dashed box. **b**, Sequence alignment of gp9 L loop, HIV fusion peptide, influenza fusion peptide and a potential hydrophobic membrane active peptide of the bacteriophage T4 tail protein gp5. Conserved residues are boxed and coloured red. Completely conserved residues are shown in white on a red background. The core region of the

HIV fusion peptide is indicated by a dashed box. **c**, Schematic diagram showing a possible mechanism for the exposure of the potential T4 hydrophobic membrane active peptide during infection. The structure of the T4 gp27 (top  $\beta$ -barrels) and gp5 (bottom) complex is shown in ribbon representation. The potential hydrophobic peptide of gp5 is coloured red. The rest of the complex structure is coloured grey. The potential hydrophobic peptide is exposed after the release of the gp5 C-terminal needle. Conformational changes of the lysozyme domain trigger the insertion of the hydrophobic peptide into the membrane.

Extended Data Table 1 | Primers used for the loop deletion mutants of gp9

Forward primer	5'-CCCATGGCATATGTACCATTATCAGGAACG-3'
Reverse primer	5'-CCGCTCGAGTCAGTGGTGGTGGTGGTGGTGCCTCAATTCATTCTCGACGC-3'
gp9Δ412-491	Forward: 5'-TAGCAATACTAAATGACCAGTTAACGAAAATGGG-3' Reverse: 5'-CCCATTTTCGTAACTGGTCATTTAGTATTGCTA-3'
gp9Δ412-486	Forward: 5'-TAGCAATACTAAATGACGCAAACATTCCGCCGCAG-3' Reverse: 5'-CTGCGGCGGAATGTTTGCGTCATTTAGTATTGCTA-3'
gp9Δ412-481	Forward: 5'-TAGCAATACTAAATGACAAGCAAGCCGATATAGC-3' Reverse: 5'-GCTATATCGGCTTGCTTGTCAATTTAGTATTGCTA-3'
gp9Δ417-491	Forward: 5'-CTATCTATCTGCTTATCAGTTAACGAAAATGG -3' Reverse: 5'-CCATTTTCGTAACTGATAAGCAGATAGATAG-3'
gp9Δ417-486	Forward: 5'-GACTATCTATCTGCTTATGCAAACATTCCGCCGCA-3' Reverse: 5'-TGCGGCGGAATGTTTGCATAAGCAGATAGATAGTC-3'
gp9Δ417-481	Forward: 5'-CTATCTATCTGCTTATAAGCAAGCCGATATAG-3' Reverse: 5'-CTATATCGGCTTGCTTATAAGCAGATAGATAG-3'
gp9Δ422-491	Forward: 5'-ATTTACAGGGCAACAAACAGTTAACGAAAATGGG-3' Reverse: 5'-CCCATTTTCGTAACTGTTTGTTGCCCTGTAAAT-3'
gp9Δ422-486	Forward: 5'-ATTTACAGGGCAACAAAGCAAACATTCCGCCGCA-3' Reverse: 5'-CTGCGGCGGAATGTTTGCTTTGTTGCCCTGTAAAT-3'
gp9Δ422-481	Forward: 5'-ATTTACAGGGCAACAAAAGCAAGCCGATATAGC-3' Reverse: 5'-GCTATATCGGCTTGCTTTTTGTTGCCCTGTAAAT-3'

Common forward and reverse primers for the overlap PCRs are listed at the top two rows of the table. Specific overlap PCR primers for each mutant are listed after the name of each deletion mutant.

Extended Data Table 2 | Crystal data collection and refinement statistics

	gp9Δ417-491	gp9Δ417-491 Hg-SAD	gp9Δ417-491	gp9 full-length
<b>Data collection</b>				
Space group	P2 <sub>1</sub> 3	P2 <sub>1</sub> 3	I222	P2 <sub>1</sub> 3
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	184.67, 184.67, 184.67	184.25, 184.25, 184.25	94.6, 135.16, 313.43	183.45, 183.45, 183.45
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97884	0.97923	0.97884	0.98
Resolution (Å)	49.35 - 2.60	50.00 - 3.50	43.32 - 2.04	45.86 - 3.50
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub> *	0.123 (0.527)	0.200 (0.591)	0.105 (0.281)	0.135 (0.456)
<i>I</i> / $\sigma$ *	16.49 (7.40)	16.43 (5.22)	17.66 (9.20)	18.36 (5.54)
Completeness (%) *	99.99 (100.00)	100.00 (100.00)	99.29 (95.11)	100.00 (100.00)
Redundancy *	11.1 (11.3)	11.5 (11.7)	6.1 (5.9)	5.4 (5.6)
<b>Refinement</b>				
Resolution (Å)	49.35 - 2.60	50.00 - 3.50	43.32 - 2.04	45.86 - 3.50
No. Reflections	64243	26595	126814	26215
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	16.64/20.07		15.31/18.32	20.07/22.86
No. Atoms				
Protein	8267		12504	9152
Ligand/Ion	14		21	
Water	327		1312	
B-factors				
Protein	47.6		27.3	73.0
Ligand/Ion	40.5		36.7	
Water	51.8		48.9	
R.m.s deviations				
Bond lengths	0.009		0.008	0.005
Bond angles (°)	1.18		1.05	1.26

\*Highest resolution shell is shown in parentheses.

# AMPK–SKP2–CARM1 signalling cascade in transcriptional regulation of autophagy

Hi-Jai R. Shin<sup>1\*</sup>, Hyunkyung Kim<sup>1\*</sup>, Sungryong Oh<sup>1</sup>, Jun-Gi Lee<sup>1</sup>, Minjung Kee<sup>1</sup>, Hyun-Jeong Ko<sup>2</sup>, Mi-Na Kweon<sup>3</sup>, Kyoung-Jae Won<sup>4</sup> & Sung Hee Baek<sup>1</sup>

**Autophagy is a highly conserved self-digestion process, which is essential for maintaining homeostasis and viability in response to nutrient starvation<sup>1–4</sup>. Although the components of autophagy in the cytoplasm have been well studied<sup>5,6</sup>, the molecular basis for the transcriptional and epigenetic regulation of autophagy is poorly understood. Here we identify co-activator-associated arginine methyltransferase 1 (CARM1) as a crucial component of autophagy in mammals. Notably, CARM1 stability is regulated by the SKP2-containing SCF (SKP1-cullin1-F-box protein) E3 ubiquitin ligase in the nucleus, but not in the cytoplasm, under nutrient-rich conditions. Furthermore, we show that nutrient starvation results in AMP-activated protein kinase (AMPK)-dependent phosphorylation of FOXO3a in the nucleus, which in turn transcriptionally represses SKP2. This repression leads to increased levels of CARM1 protein and subsequent increases in histone H3 Arg17 dimethylation. Genome-wide analyses reveal that CARM1 exerts transcriptional co-activator function on autophagy-related and lysosomal genes through transcription factor EB (TFEB). Our findings demonstrate that CARM1-dependent histone arginine methylation is a crucial nuclear event in autophagy, and identify a new signalling axis of AMPK–SKP2–CARM1 in the regulation of autophagy induction after nutrient starvation.**

To explore the importance of nuclear events in autophagy, we proposed that specific histone marks are involved in the epigenetic and transcriptional regulation of autophagy in the nucleus leading to the fine-tuning of the autophagy process. We induced autophagy in mouse embryonic fibroblasts (MEFs) by glucose starvation, and sought to identify altered specific histone marks. We observed an increase in histone H3 Arg17 dimethylation (H3R17me2) levels in response to glucose starvation (Fig. 1a), which also occurred when autophagy was triggered by amino acid starvation or rapamycin (Extended Data Fig. 1a). Notably, nutrient starvation resulted in increased levels of CARM1 protein (Fig. 1b and Extended Data Fig. 1b).

To examine whether CARM1 induction and subsequent increases in H3R17me2 are related to autophagy occurrence, we analysed the conversion of non-lipidated LC3-I to lipidated LC3-II, as a common marker of autophagic activity<sup>7</sup>. The increase in CARM1 was associated with an increase in LC3-II (Fig. 1c and Extended Data Fig. 1c, d). To confirm that the decrease in LC3-II reflects decreases in functional autophagic degradation, autophagic flux was also analysed using the levels of p62 (also known as SQSTM1)<sup>8,9</sup>. Glucose starvation induced p62 degradation and LC3-II accumulation in wild-type MEFs but not in *Carm1* knockout and knock-in MEFs expressing the enzymatic activity-deficient mutant (Fig. 1c).

To evaluate the role of CARM1 in the autophagic process, the formation of green fluorescent protein (GFP)-tagged LC3-positive autophagosome was examined. The increase in GFP–LC3 punctate cells was

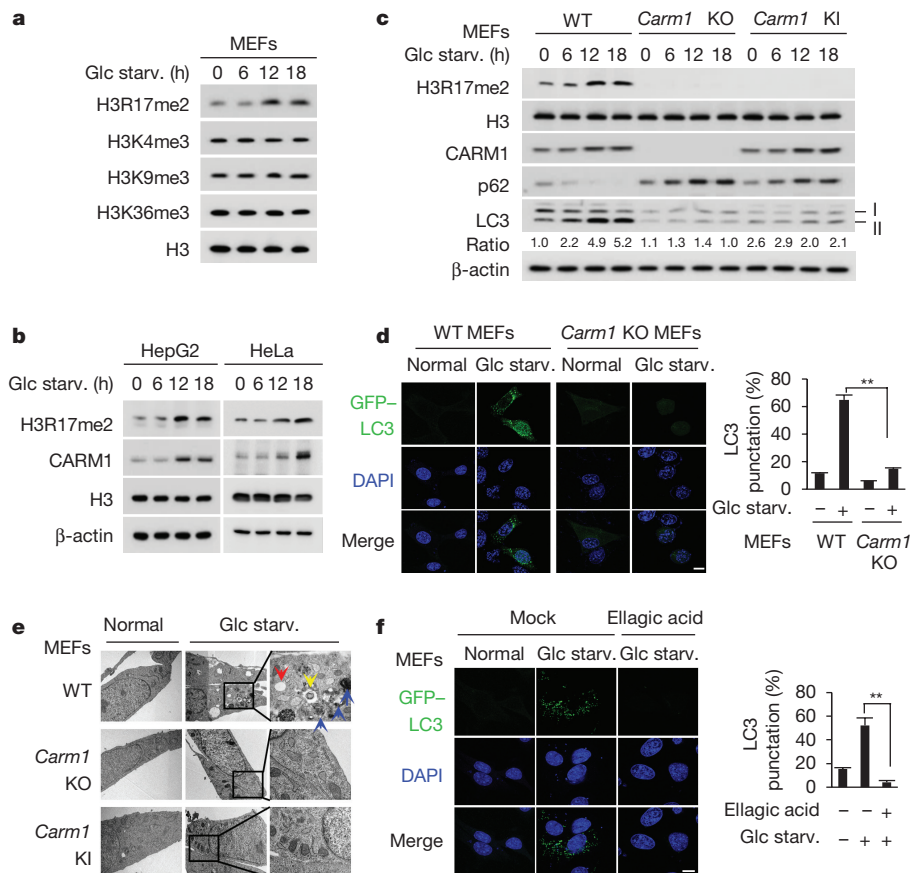
notably attenuated in *Carm1* knockout compared to wild-type MEFs (Fig. 1d and Extended Data Fig. 1e). Transmission electron microscopy (TEM) further showed an increase in the number of autophagic vesicles in wild-type MEFs, but not in *Carm1* knockout and knock-in MEFs (Fig. 1e). We performed LC3 flux analysis using bafilomycin A1, an inhibitor of the late phase of autophagy. Defects in autophagic flux caused by the loss of CARM1 were confirmed by immunoblot analysis (Extended Data Fig. 2a, b) and imaging experiments using mCherry–GFP–LC3, which provides a simultaneous readout of autophagosome formation and maturation (Extended Data Fig. 2c). In addition, ellagic acid, a naturally occurring polyphenol reported to selectively inhibit H3R17me2 (ref. 10), greatly compromised the autophagic process (Fig. 1f and Extended Data Fig. 2d–f).

Next, we examined how CARM1 induction is regulated after glucose starvation. We found that CARM1 protein levels were increased only in the nucleus after glucose starvation (Fig. 2a, left). Treatment of MG132, a 26S proteasome inhibitor, inhibited nuclear CARM1 degradation (Fig. 2a, right). Glucose starvation markedly reduced the ubiquitination of CARM1 in the nucleus, whereas CARM1(K471R) failed to be ubiquitinated, indicating that K471 is the ubiquitination-targeting site (Fig. 2b and Extended Data Fig. 3a). We then sought to identify the E3 ubiquitin ligase responsible for CARM1 ubiquitination. Notably, SKP2, an F-box protein of the SCF E3 ubiquitin ligase complex, was identified as a CARM1-binding protein along with cullin 1 (CUL1) (Fig. 2c and Supplementary Table 1). CARM1 exhibited specific binding to SKP2 (Fig. 2d) and CUL1 (Extended Data Fig. 3b).

Since CARM1 is stabilized after glucose starvation and possibly ubiquitinated by the SKP2-containing E3 ligase complex under nutrient-rich condition, we checked for changes in SKP2 protein levels. A reduction in SKP2 and an increase in CARM1 protein levels were observed in glucose-starved cells (Fig. 2e). Decreased levels of SKP2 resulted in the stabilization of other known SKP2–SCF substrates (Extended Data Fig. 3c). Furthermore, SKP2 knockdown attenuated CARM1 ubiquitination in the nucleus (Fig. 2f) and markedly increased the half-life of CARM1 (Extended Data Fig. 3d). By contrast, overexpression of wild-type SKP2, but not the SKP2ΔF mutant that is not able to form a SKP2–SCF complex<sup>11</sup>, decreased the half-life and protein levels of CARM1 in cells deprived of glucose (Fig. 2g, h and Extended Data Fig. 3e). We speculate that exclusive nuclear localization of SKP2 results in selective CARM1 ubiquitination in the nucleus. As a result of SKP2 downregulation, the interaction between CUL1 and CARM1 significantly decreased after glucose starvation (Extended Data Fig. 4a). Also, as a component of the SCF complex, CUL1 regulated CARM1 protein levels (Extended Data Fig. 4b–e). Collectively, these data indicate that the SKP2-containing SCF E3 ligase complex is responsible for CARM1 degradation in the nucleus under nutrient-rich conditions (Fig. 2i).

<sup>1</sup>Creative Research Initiatives Center for Chromatin Dynamics, School of Biological Sciences, Seoul National University, Seoul, South Korea. <sup>2</sup>Laboratory of Microbiology and Immunology, College of Pharmacy, Kangwon National University, Chuncheon, South Korea. <sup>3</sup>Mucosal Immunology Laboratory, Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea. <sup>4</sup>The Institute for Diabetes, Obesity, and Metabolism, Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

\*These authors contributed equally to this work.



**Figure 1 | Increased H3R17 dimethylation by CARM1 is critical for proper autophagy.** **a**, **b**, Immunoblot analysis of various histone marks and CARM1 in response to glucose starvation (Glc starv.). **c**, Wild-type (WT), *Carm1* knockout (KO) or knock-in (KI) MEFs were subject to immunoblot analysis. The LC3-II/LC3-I ratio is indicated. **d**, Representative confocal images of GFP-LC3 puncta formation. Graph shows quantification of LC3-positive punctate cells (right).

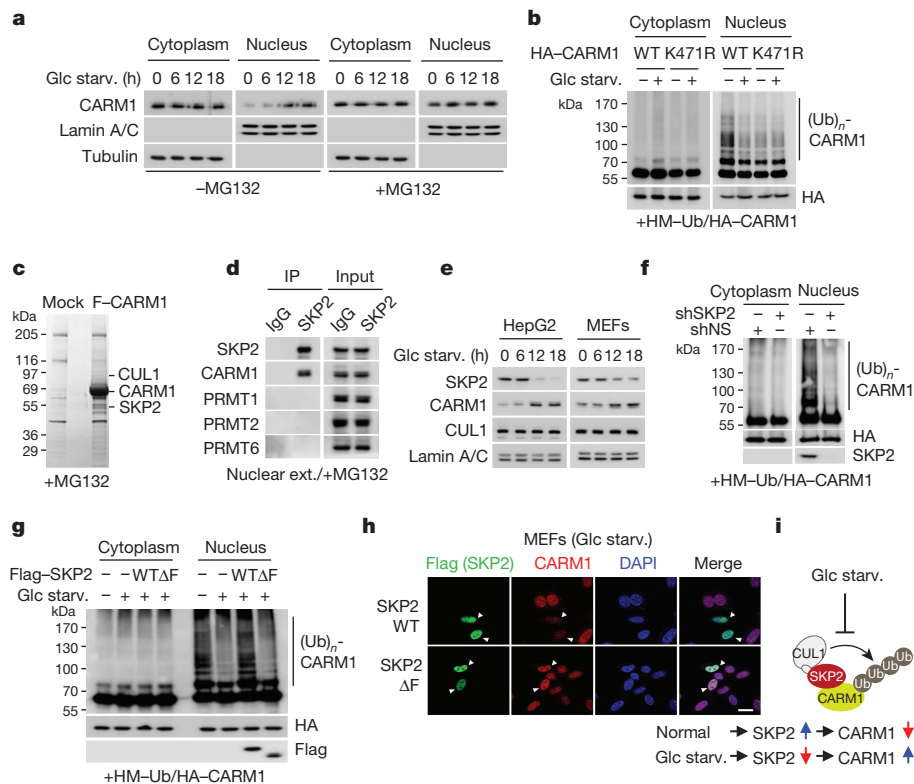
It has been shown that AMPK is activated during glucose starvation and leads to starvation-induced autophagy<sup>12–14</sup>. As the role of nuclear AMPK in autophagy outcome has not been defined thus far, we aimed to examine whether AMPK is involved in the transcriptional regulation of autophagy. We found that AMPK $\alpha$ 2 and phosphorylated AMPK, the activated form of AMPK, increased in the nucleus after glucose starvation (Fig. 3a). Increased AMPK $\alpha$ 2 resulted from transcription induction rather than post-translational regulation (Extended Data Fig. 5a–c). AMPK $\alpha$ 2 has been shown to be preferentially expressed in the nucleus<sup>15</sup>, suggesting that it might perform distinct roles in the nucleus. AMPK failed to directly bind or phosphorylate CARM1 and SKP2 (Extended Data Fig. 5d, e). However, AMPK activation by aminoimidazole carboxamide ribonucleotide (AICAR) and phenformin resulted in the increase of CARM1 and reduction of SKP2 (Extended data Fig. 5f), and this was compromised when AMPK activity was blocked by compound C (Extended data Fig. 5g).

We then used wild-type and *Ampk* $\alpha$ 2 (also known as *Prkaa1* and *Prkaa2*) double knockout (DKO) MEFs to check for the expression of CARM1 and SKP2. In the nucleus, CARM1 induction and SKP2 reduction after glucose starvation were abrogated in *Ampk* DKO MEFs (Fig. 3b). The half-life of CARM1 in the nucleus was decreased in *Ampk* DKO MEFs (Extended Data Fig. 5h). Introduction of wild-type AMPK $\alpha$ 2, but not the dominant-negative form, in *Ampk* DKO MEFs resulted in a recovered expression pattern of SKP2 and CARM1, similar to wild-type MEFs (Extended Data Fig. 5i). SKP2 depletion in *Ampk* DKO MEFs led to increased CARM1 protein levels, indicating that the reduction of CARM1 in *Ampk* DKO MEFs is

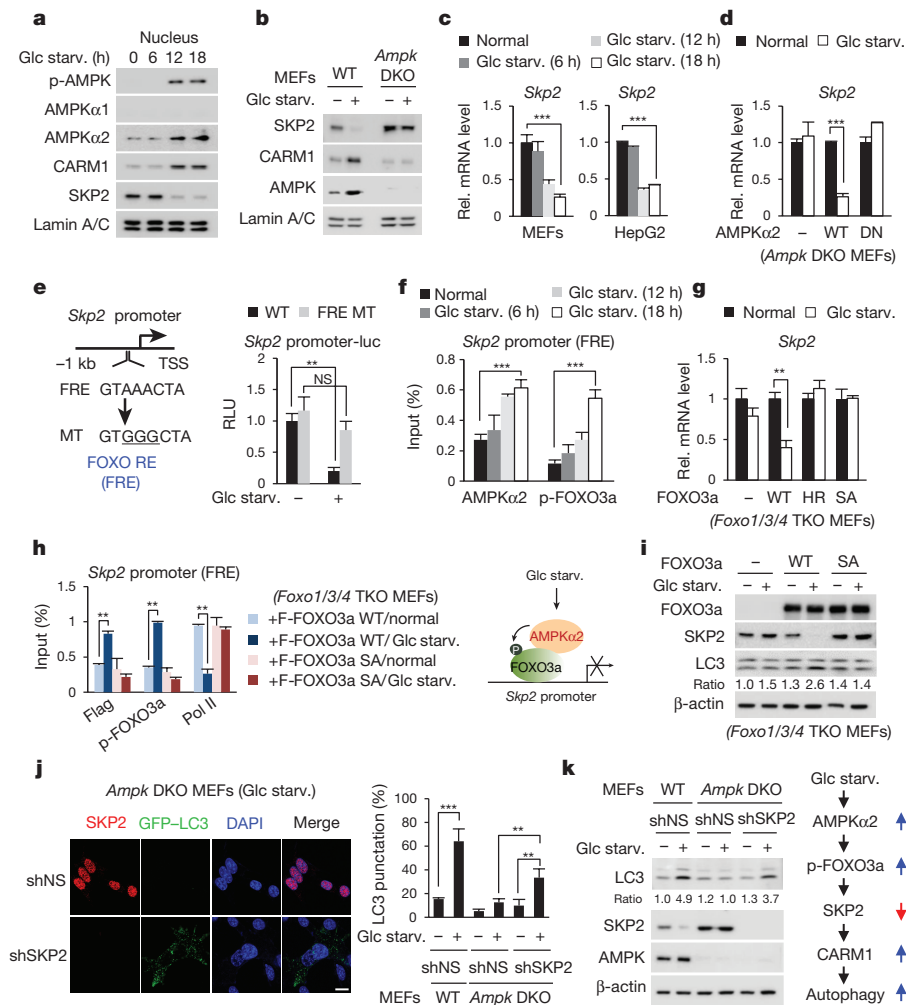
mediated by SKP2 (Extended Data Fig. 5j). Furthermore, since binding of CARM1 to CUL1 is mediated by SKP2, the CARM1–CUL1 interaction was maintained upon glucose starvation in *Ampk* DKO MEFs (Extended Data Fig. 5k).

Reduction of SKP2 expression after glucose starvation is not mediated by proteasomal degradation (Extended Data Fig. 5l), but instead regulated at the transcription level (Fig. 3c). Glucose starvation failed to decrease *Skp2* mRNA levels in *Ampk* DKO MEFs, but reconstitution of wild-type AMPK $\alpha$ 2 restored the reduction in *Skp2* mRNA (Fig. 3d). Therefore, we were prompted to search for a possible regulatory mechanism of SKP2 downregulation by AMPK $\alpha$ 2. Recent studies have emphasized the AMPK–FOXO axis as a highly conserved nutrient-sensing pathway crucial for cellular and organismal homeostasis<sup>16</sup>. AMPK directly phosphorylates FOXO3a and regulates FOXO3a transcriptional activity<sup>17</sup>. Although mainly known as a transcriptional activator, FOXO also functions as a transcriptional repressor<sup>18–21</sup>. *Skp2* promoter analysis revealed a highly conserved FOXO response element (FRE) (Fig. 3e). We proposed that FOXO might function as a transcriptional repressor of SKP2 and performed luciferase reporter assay driven by the *Skp2* promoter. Glucose starvation attenuated *Skp2* promoter luciferase activity, but not the *Skp2* promoter containing an FRE mutation (Fig. 3e). *Skp2* mRNA levels failed to decrease in *Foxo1/3/4* triple knockout (TKO) MEFs (Extended Data Fig. 5m), indicating that FOXO is a crucial transcription factor in the repression of SKP2.

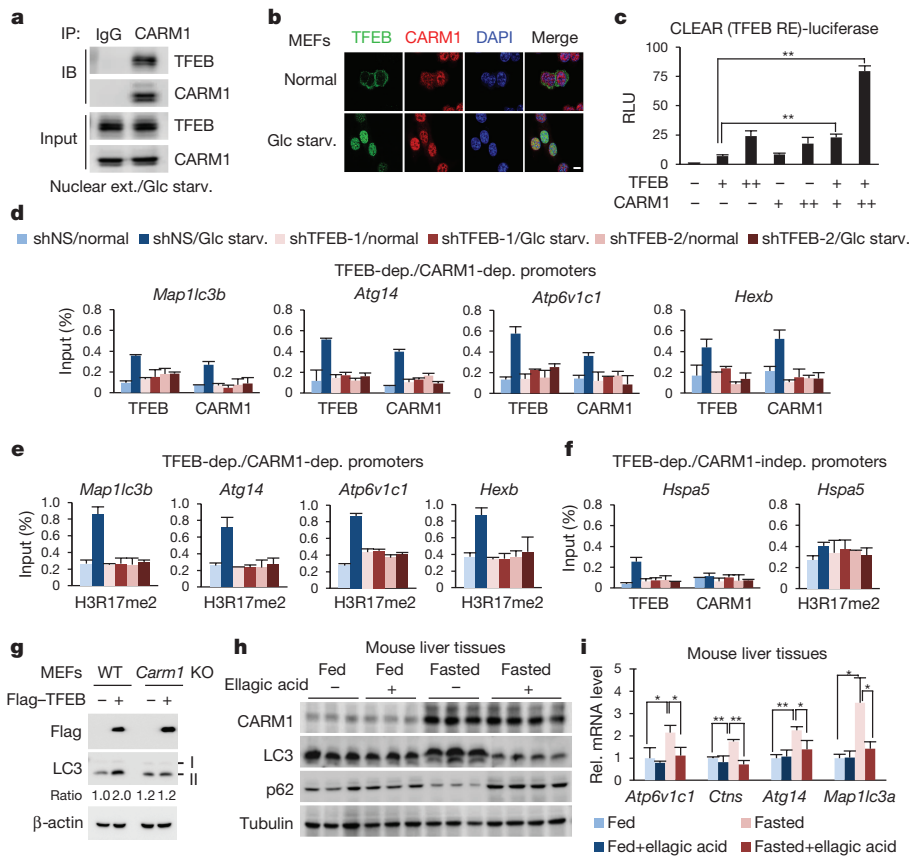
Glucose starvation resulted in AMPK-dependent FOXO3a phosphorylation (Extended Data Fig. 5n). In addition, AMPK $\alpha$ 2 and phosphorylated FOXO3a were co-recruited to the *Skp2* promoter



**Figure 2 | CARM1 is degraded by the SKP2-containing SCF E3 ligase in the nucleus under nutrient-rich conditions.** **a**, MEFs were deprived of glucose in the absence (left) or presence (right) of MG132 ( $5 \mu\text{g ml}^{-1}$ ) and subject to immunoblotting. **b**, *In vivo* ubiquitination assay of wild-type CARM1 or ubiquitination-defective K471R mutant CARM1. HA, haemagglutinin; HM, HisMax tag. **c**, Identification of CARM1-interacting proteins. F-CARM1 denotes Flag-tagged CARM1 construct. **d**, Interactions between SKP2 and protein arginine methyltransferases (PRMTs) were analysed. **e**, Glucose-starved cells were subjected to immunoblotting. **f**, *In vivo* ubiquitination assay of CARM1. **g**, Representative confocal images. Scale bar,  $20 \mu\text{m}$ . **h**, Schematic of SKP2-containing SCF E3 ligase-dependent degradation of CARM1.



**Figure 3 | Decrease in SKP2 after glucose starvation is AMPK dependent.** **a**, MEFs deprived of glucose were analysed with the indicated antibodies. **b**, Nuclear fractions from wild-type and *Ampk* double knockout (DKO) MEFs were subjected to immunoblotting. **c**, *qRT-PCR* of *Skp2*. DN, dominant negative. **d**, Left, schematic of *Skp2* promoter. Right, luciferase activities of wild-type *Skp2* or FOXO response element (FRE) mutant promoter were measured. MT, mutant; RE, response element; RLU, relative light units; TSS, transcription start site. **e**, ChIP assays of the *Skp2* promoter. **f**, *Skp2* mRNA levels were analysed in *Foxo1/3/4* triple knockout (TKO) MEFs. HR, H212R mutant; SA, sextuple T179A/S399A/S413A/S555A/S588A/S626A mutant. **g**, Left, ChIP assays of the *Skp2* promoter. Right, schematic of SKP2 regulation by the AMPK-FOXO axis. **h**, Bars, mean  $\pm$  s.e.m.;  $n = 3$ . NS, not significant. **i**,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$  (one-tailed *t*-test) (**c-h**). **j**, Representative confocal images of GFP-LC3 puncta formation. shSKP2, short-hairpin RNA (shRNA) against SKP2; shNS, nonspecific shRNA. Scale bar,  $20 \mu\text{m}$ . Bars, mean  $\pm$  s.e.m.;  $n = 5$ , with over 80 cells;  $^{**}P < 0.01$ ,  $^{***}P < 0.001$  (one-tailed *t*-test). **k**, Left, immunoblot analysis from whole-cell extracts. Right, schematic of the AMPK-SKP2-CARM1 signalling cascade in autophagy.



**Figure 4 | CARM1 exerts a transcriptional co-activator function on autophagy-related and lysosomal genes through TFEB.** **a**, Binding between CARM1 and TFEB. **b**, Representative confocal images. Scale bar, 10  $\mu$ m. **c**, 2 $\times$  CLEAR (TFEB RE)-luciferase reporter assays. Bars, mean  $\pm$  s.e.m.;  $n = 3$ . **d**, **e**, ChIP assays on TFEB-dependent, CARM1-dependent (**d**) or CARM1-independent (**e**) promoters after knockdown of TFEB. Bars, mean  $\pm$  s.e.m.;  $n = 3$ .

upon glucose starvation (Fig. 3f). The recruitment of phosphorylated FOXO3a accompanied by a decrease in RNA polymerase II was also observed in *Ampk* DKO MEFs reconstituted with wild-type AMPK $\alpha$ 2 (Extended Data Fig. 5o). Notably, reconstitution of wild-type FOXO3a in *Foxo1/3/4* TKO MEFs significantly reduced the *Skp2* mRNA level, but neither the FOXO3a(H212R) DNA-binding mutant<sup>22</sup> nor the FOXO3a sextuple SA mutant, which is not phosphorylated by AMPK<sup>17</sup>, reduced *Skp2* mRNA levels (Fig. 3g). Furthermore, after glucose starvation, phosphorylated FOXO3a, but not the FOXO3a SA mutant, was recruited to the *Skp2* promoter (Fig. 3h), indicating that AMPK-dependent FOXO3a phosphorylation is crucial for the recruitment of FOXO3a at the *Skp2* promoter. SKP2 expression failed to decrease and autophagy occurrence was impaired in FOXO3a SA mutant-reconstituted *Foxo1/3/4* TKO MEFs (Fig. 3i).

We observed a marked increase in autophagy occurrence in *Ampk* DKO MEFs after SKP2 knockdown (Fig. 3j, k). We also tested whether CARM1 overexpression could restore autophagy in *Ampk* DKO MEFs. Introduction of wild-type or K471R mutant CARM1 restored the number of GFP-LC3 punctate cells, whereas enzymatic-dead mutant CARM1(R169A) failed to do so (Extended Data Fig. 5p). Collectively, we found a signalling axis in autophagy induction in which glucose starvation activates AMPK $\alpha$ 2 in the nucleus, leading to transcriptional repression of *Skp2* via FOXO3a phosphorylation. Reduction of SKP2 expression in turn leads to increased levels of CARM1.

To gain insight into the role of CARM1 in transcriptional regulation of autophagy, we performed RNA-sequencing (RNA-seq) in wild-type and *Carm1* knockout MEFs after glucose starvation (Extended Data Fig. 6a, b). Using a comprehensive list of known autophagy-related

and lysosomal genes (Supplementary Table 2), we found that potential CARM1 target genes (cluster 1) are significantly enriched for autophagy-related and lysosomal genes (Extended Data Fig. 6c). Transcription factor motif analysis indicated TFEB as a putative major transcription factor for CARM1 (Extended Data Fig. 6d). We validated CARM1 dependency of the autophagy-related and lysosomal genes by quantitative reverse transcription PCR (qRT-PCR) (Extended Data Fig. 6e). Furthermore, we performed chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) of H3R17me2 in wild-type MEFs after glucose starvation and observed enriched H3R17me2 as well as activating H3K4me3 signals at active promoters (Extended Data Fig. 6f–h and Supplementary Table 3).

**g**, Wild-type and *Carm1* knockout MEFs transfected with Flag-TFEB were subject to immunoblot analysis. **h**, Liver tissues from fed or fasted mice treated with vehicle or ellagic acid were subjected to immunoblot analysis ( $n = 3$  per group). **i**, Expression of autophagy-related genes and lysosomal genes in wild-type mouse livers. Bars, mean  $\pm$  s.e.m.;  $n = 3$  per group.  $^*P < 0.05$ ,  $^{**}P < 0.01$  (two-tailed  $t$ -test).

TFEB functions as a master regulator of lysosomal biogenesis and autophagy<sup>23–25</sup>. After glucose starvation, CARM1 and TFEB exhibited mutual binding in the nucleus (Fig. 4a, b and Extended Data Fig. 7a). The binding of CARM1 to TFEB was not affected by AMPK (Extended Data Fig. 7b). CARM1 binds to the transcriptional activation domain of TFEB, whereas TFEB binds to the methyltransferase domain of CARM1 (Extended Data Fig. 7c, d). Although CARM1 also interacts with TFE3, TFEB knockdown, but not TFE3 knockdown, markedly altered the transcription induction of various target genes (Extended Data Fig. 7e–h).

Introduction of TFEB increased CLEAR-element-containing luciferase reporter activity and overexpression of CARM1 further enhanced its activity (Fig. 4c). To examine whether CARM1-dependent target genes are regulated by TFEB, we searched for putative CLEAR motif (Supplementary Table 2) and performed ChIP assays. Knockdown of TFEB abolished the recruitment of CARM1 to its target promoters,

subsequently leading to the failure of H3R17me2 induction (Fig. 4d, e). CARM1 recruitment was not observed on CARM1-independent promoters (Fig. 4f). Conversely, a subset of TFEB target genes failed to increase upon glucose starvation after CARM1 knockdown (Extended Data Fig. 7i). CARM1 depletion was accompanied by a reduction in H3R17me2 on TFEB-dependent, CARM1-dependent target promoters, with little or no effect on TFEB recruitment (Extended Data Fig. 8a, b). Immunoblot analysis confirmed several key autophagy regulators that are transcriptionally regulated by CARM1 were induced by glucose starvation in wild-type MEFs, but not in *Carm1* knockdown or knockout MEFs (Extended Data Fig. 8c). Furthermore, a two-step ChIP assay confirmed the recruitment of CARM1 at TFEB-bound genes (Extended Data Fig. 8d).

Previous studies reported that overexpression of TFEB induces autophagy<sup>24</sup>. However, introduction of TFEB in *Carm1* knockout MEFs failed to increase the formation of autophagosomes and levels of LC3-II (Fig. 4g and Extended Data Fig. 8e). As CARM1 fails to increase upon glucose starvation in *Ampk* DKO MEFs, TFEB-dependent, CARM1-dependent target gene expression and induction of H3R17me2 were dampened in *Ampk* DKO MEFs (Extended Data Fig. 9a, b). However, SKP2 knockdown significantly increased the mRNA levels of CARM1 target genes (Extended Data Fig. 9c), indicating that partial recovery of autophagy in *Ampk* DKO MEFs by SKP2 knockdown is due to transcriptional activation of autophagy-related and lysosomal genes. Collectively, these data indicate CARM1 as a crucial co-activator of TFEB.

To examine whether CARM1 and subsequent H3R17me2 are important for autophagy occurrence *in vivo*, we analysed hepatic autophagy in wild-type mice. Livers of fasted mice showed a marked increase in CARM1 levels, as well as an increase in LC3 conversion. However, LC3 conversion was greatly attenuated in fasted livers of mice treated with ellagic acid (Fig. 4h). Furthermore, the mRNA expression of various CARM1-dependent autophagy-related and lysosomal genes failed to increase (Fig. 4i). Ellagic acid treatment inhibited the induction of a subset of autophagy-related and lysosomal genes regulated by CARM1, and blocked the recruitment of CARM1, but not TFEB, along with reduced H3R17me2 levels at CARM1-dependent promoters (Extended Data Fig. 9d–f). Given that the inhibition of H3R17me2 by ellagic acid almost completely blocks CARM1-induced autophagy occurrence, ellagic acid might have the potential to be developed as a therapeutic agent in autophagy-related diseases.

Here, we provide a link between energy sensing, chromatin modifications and transcriptional and epigenetic regulation of autophagy (Extended Data Fig. 10). Although our current work is focused on CARM1 stabilization, we speculate that this type of regulation in the nucleus might be an efficient way to regulate target gene expression, and could be a prototype of protein stabilization for histone modifiers. In addition, our data indicate that when glucose starvation persists and transcription of various autophagy-related genes is needed to sustain autophagy, AMPK accumulates in the nucleus and actively controls transcription. Our findings shed light on the potential therapeutic targeting of a new signalling axis of AMPK–SKP2–CARM1 in autophagy-related diseases.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 September 2015; accepted 14 April 2016.**

**Published online 15 June 2016.**

- Yang, Z. & Klionsky, D. J. Eaten alive: a history of macroautophagy. *Nat. Cell Biol.* **12**, 814–822 (2010).
- Mizushima, N., Levine, B., Cuervo, A. M. & Klionsky, D. J. Autophagy fights disease through cellular self-digestion. *Nature* **451**, 1069–1075 (2008).

- Rabinowitz, J. D. & White, E. Autophagy and metabolism. *Science* **330**, 1344–1348 (2010).
- Choi, A. M., Ryter, S. W. & Levine, B. Autophagy in human health and disease. *N. Engl. J. Med.* **368**, 651–662 (2013).
- Mizushima, N. Autophagy: process and function. *Genes Dev.* **21**, 2861–2873 (2007).
- Klionsky, D. J. Autophagy: from phenomenology to molecular understanding in less than a decade. *Nat. Rev. Mol. Cell Biol.* **8**, 931–937 (2007).
- Mizushima, N. & Yoshimori, T. How to interpret LC3 immunoblotting. *Autophagy* **3**, 542–545 (2007).
- Mizushima, N., Yoshimori, T. & Levine, B. Methods in mammalian autophagy research. *Cell* **140**, 313–326 (2010).
- Bjørkøy, G. *et al.* Monitoring autophagic degradation of p62/SQSTM1. *Methods Enzymol.* **452**, 181–197 (2009).
- Selvi, B. R. *et al.* Identification of a novel inhibitor of coactivator-associated arginine methyltransferase 1 (CARM1)-mediated methylation of histone H3 Arg-17. *J. Biol. Chem.* **285**, 7143–7152 (2010).
- Carrano, A. C., Eytan, E., Hershko, A. & Pagano, M. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nat. Cell Biol.* **1**, 193–199 (1999).
- Hardie, D. G. AMPK and autophagy get connected. *EMBO J.* **30**, 634–635 (2011).
- Mihaylova, M. M. & Shaw, R. J. The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. *Nat. Cell Biol.* **13**, 1016–1023 (2011).
- Inoki, K., Kim, J. & Guan, K.-L. AMPK and mTOR in cellular energy homeostasis and drug targets. *Annu. Rev. Pharmacol. Toxicol.* **52**, 381–400 (2012).
- Salt, I. *et al.* AMP-activated protein kinase: greater AMP dependence, and preferential nuclear localization, of complexes containing the  $\alpha 2$  isoform. *Biochem. J.* **334**, 177–187 (1998).
- Eijkelenboom, A. & Burgering, B. M. FOXOs: signalling integrators for homeostasis maintenance. *Nat. Rev. Mol. Cell Biol.* **14**, 83–97 (2013).
- Greer, E. L. *et al.* The energy sensor AMP-activated protein kinase directly regulates the mammalian FOXO3 transcription factor. *J. Biol. Chem.* **282**, 30107–30119 (2007).
- Potente, M. *et al.* Involvement of Foxo transcription factors in angiogenesis and postnatal neovascularization. *J. Clin. Invest.* **115**, 2382–2392 (2005).
- Wang, K. & Li, P.-F. Foxo3a regulates apoptosis by negatively targeting miR-21. *J. Biol. Chem.* **285**, 16958–16966 (2010).
- Yang, Y.-C. *et al.* DNMT3B overexpression by deregulation of FOXO3a-mediated transcription repression and MDM2 overexpression in lung cancer. *J. Thorac. Oncol.* **9**, 1305–1315 (2014).
- Lam, E. W.-F., Brosens, J. J., Gomes, A. R. & Koo, C.-Y. Forkhead box proteins: tuning forks for transcriptional harmony. *Nat. Rev. Cancer* **13**, 482–495 (2013).
- Tsai, K.-L. *et al.* Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res.* **35**, 6984–6994 (2007).
- Sardiello, M. *et al.* A gene network regulating lysosomal biogenesis and function. *Science* **325**, 473–477 (2009).
- Settembre, C. *et al.* TFEB links autophagy to lysosomal biogenesis. *Science* **332**, 1429–1433 (2011).
- Settembre, C. & Medina, D. L. TFEB and the CLEAR network. *Methods Cell Biol.* **126**, 45–62 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank members of the Chromatin Dynamics Research Center for technical assistance and discussions and J. Kim and J. Chung for valuable reagents and discussions. We thank Y. S. Yu for illustrations. The TEM data were analysed in the Korean Basic Science Institute. *Carm1* knockout and knock-in MEFs were provided by M. T. Bedford. *Ampk* DKO MEFs was a gift from B. Viollet, and *Foxo1.3.4<sup>+/+</sup>* MEFs were a gift from R. DePinho and J.-H. Paik. This work was supported by Creative Research Initiatives Program (Research Center for Chromatin Dynamics, 2009-0081563) to S.H.B.; the National Junior Research Fellowship (NRF-2011-A01496-0001806) to H.-J.R.S.; the Basic Science Research Program (NRF-2014R1A6A3A0405 7910) to H.K. from the National Research Foundation (NRF) grant funded by the South Korean government (MSIP); NIH grant (R01DK106027) to K.-J.W.

**Author Contributions** H.-J.R.S., H.K., S.O., J.-G.L. and M.K. performed the cell biology and biochemistry experiments; H.-J.K. and M.-N.K. provided TEM analysis and critical comments; H.-J.R.S. and K.J.W. performed RNA and ChIP-seq preparation and systemic analysis; H.-J.R.S., H.K., K.J.W. and S.H.B. organized and analysed the data; H.-J.R.S., K.J.W. and S.H.B. wrote the manuscript.

**Author Information** The RNA-seq and H3R17me2 ChIP-seq data sets have been deposited in the NCBI Gene Expression Omnibus (GEO) database under the accession number GSE72901. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.H.B. (sbaek@snu.ac.kr).

## METHODS

**Antibodies and reagents.** The following commercially available antibodies were used: anti-AMPK $\alpha$ 1 (ab110036), anti-AMPK $\alpha$ 2 (ab3760), anti-ATG14 (ab173943), anti-FOXO3a (ab12162), anti-histone H3 (ab1791), anti-H3R17me2 (ab8284), anti-H3K4me3 (ab8580), anti-H3K9me3 (ab8898), anti-H3K36me3 (ab9050), anti-PI3K class 3 (ab124905), and anti-TFEB (ab2636) antibodies were purchased from Abcam. Anti-AMPK (2532), anti-ATG12 (4180), anti-CARM1 (3379 for immunoblotting, 12495 for immunoprecipitation and ChIP), anti-LC3 (2775), anti-phospho-AMPK $\alpha$  T172 (2535), anti-phospho-FOXO3a S413 (8174), anti-SQSTM1/p62 (5114), and anti-TFE3 (14779) antibodies were from Cell Signaling Technology. Anti-SKP2 (sc-7164), anti-CUL1 (sc-17775), anti-tubulin (sc-8035), and anti-Lamin A/C (sc-6215) were from Santa Cruz Biotechnology. Anti-Flag (F3165), anti-ULK1 (A7481) and anti- $\beta$ -actin (A1978) antibodies were from Sigma, anti-HA antibody (MMS-101R) from Covance, and anti-tubulin antibody (LF-PA0146A) from Abfrontier. The following chemicals were used in this study: rapamycin (R-5000) was purchased from LC laboratories, cycloheximide (C4859), AICAR (A9978) and phenformin (P7045) from Sigma, bafilomycin A1 (11038) and ellagic acid from Cayman (10569), compound C from Calbiochem (171260), and MG132 (M-1157) from A.G. Scientific.

**Cell culture and generation of shRNA knockdown cells.** HEK293T, HeLa and HepG2 cells, and wild-type, *Carm1* knockout, *Carm1* knock-in, *Ampk* DKO and *Foxo1/3/4*<sup>+/−</sup> MEFs were cultured at 37 °C in DMEM containing 10% fetal bovine serum (FBS) and antibiotics in a humidified incubator with 5% CO<sub>2</sub>. All cell lines used in the study were regularly tested for mycoplasma contamination. For glucose starvation, cells were washed with PBS, then incubated with glucose-free DMEM supplemented with 10% dialysed FBS. Transfection was performed with Turbofect (Fermentas) or Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol. To generate knockdown cells, lentiviral shRNA constructs were first transfected along with viral packaging plasmids (psPAX2 and pMD2.G) into HEK293T cells. Three days after transfection, viral supernatant was filtered through 0.45- $\mu$ m filter and infected into targeting cells. Infected cells were then selected with 5  $\mu$ M puromycin. The targeting sequences of shRNAs are as follows. mCARM1-1; 5'-TCAGGGACATGTCTGCTTATT-3', mCARM1-2; 5'-GCCTGAGCAAGTGGACATTAT-3', mTFE3-1; 5'-GTG GATTACATCCGCAATTA-3', mTFE3-2; 5'-TGTGGATTACATCCGCA AATT-3', mTFE3-3; 5'-GC AGGCTGTCATGCATTATAT-3', mTFE3-4; 5'-CC AAGAAGACTCTGGACTTAA-3', mSKP2; 5'-GCAAGACTTCTGAAGCTG CTAT-3', hCUL1-1; 5'-GATTGTGATGGATGAGAGTG TA-3', hCUL1-2; 5'-CC CGCAGCAAATAGTTTCATGT-3', hSKP2-1; 5'-TTCCGCTGCCACGATCA TTT-3', hSKP2-2; 5'-AGTCGGTGCTATGATATAATA-3'.

**Animal studies.** All animal studies and procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Seoul National University. Eight-to-ten-week-old male wild-type C57BL/6J mice were injected with vehicle (PEG400) or ellagic acid (10 mg kg<sup>−1</sup> day<sup>−1</sup>) intraperitoneally for four consecutive days. Mice were then fed *ad libitum* or fasted for 24 h. Liver tissues were collected after mice were euthanized. Sample sizes were at least  $n = 3$  to allow for statistical analysis.

**Whole-cell lysate preparation and subcellular fractionation.** All cells were briefly rinsed with ice-cold PBS before collection. For whole-cell lysates, the cells were resuspended in RIPA buffer (150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl (pH 7.5), and 2 mM EDTA (pH 8.0)) supplemented with protease inhibitors and sonicated using a Branson Sonifier 450 at output 3 and a duty cycle of 30 for five pulses. For cytosolic and nuclear fractions, cells were lysed in harvest buffer (10 mM HEPES (pH 7.9), 50 mM NaCl, 0.5 M sucrose, 0.1 mM EDTA, 0.5% Triton X-100 and freshly added DTT, PMSF and protease inhibitors), incubated on ice for 5 min and spun at 120g for 10 min at 4 °C. The supernatant (cytosolic fraction) was removed to a separate tube. The nuclear pellet was rinsed twice with 500  $\mu$ l of buffer A (10 mM HEPES (pH 7.9), 10 mM KCl, 0.1 mM EDTA, and 0.1 mM EGTA) and spun down at 120g for 10 min at 4 °C. The supernatant was discarded and the pellet (nuclear fraction) was resuspended in RIPA buffer and sonicated as for the whole-cell lysates. All lysates were quantified by the Bradford method and analysed by SDS-PAGE.

**Electron microscopy.** Cells were fixed in 0.1 M sodium cacodylate containing 4% glutaraldehyde, 1% paraformaldehyde for 1 h at room temperature. After washing three times with 0.1 M sodium cacodylate, cells were dehydrated through a gradient series of ethanol, 20 min each step, starting from 50% ethanol and ending with 100% ethanol. Afterwards, cells were incubated with progressively concentrated propylene oxide dissolved in ethanol then infiltrated with increasing concentration of Eponate 812 resin. Samples were baked in a 65 °C oven overnight then sectioned using an Ultra microtome. Sections were viewed with an energy filtering TEM unit (LEO-192AB OMEGA, Carl Zeiss) at the Korean Basic Science Institute, South Korea.

**Immunofluorescence.** Immunocytochemistry was performed as previously described<sup>26</sup>. Cells grown on coverslips at a density of  $7 \times 10^4$  cells were washed three times with PBS and then fixed with 2% paraformaldehyde in PBS for 10 min at room temperature. Fixed cells were permeabilized with 0.1% Triton X-100 in PBS (PBS-T) for 10 min at room temperature. Blocking was performed with 3% bovine serum in PBS-T for 30 min. For staining, cells were incubated with antibodies for 2 h at room temperature, followed by incubation with fluorescent labelled secondary antibodies for 1 h (Invitrogen). Cells were mounted and visualized under a confocal microscope (Zeiss, LSM700). For autophagy studies, MEFs were transfected with GFP-LC3 and sub-cultured onto coverslips. The following day, cells were incubated with either complete media or glucose starvation media for 18 h. Cells were treated with rapamycin or ellagic acid for 18 h. For BiFC experiments, pHA-CARM1-VC155 and pFlag-TFEB-VN173 constructs were used.

**Ubiquitination assay.** Ubiquitination assay was performed as previously described<sup>27</sup>. Cells were transfected with combinations of plasmids including HisMax-tagged ubiquitin. After incubation for 48 h, cells were treated with 5  $\mu$ M ml<sup>−1</sup> of MG132 for 4 h, lysed in buffer A (6 M guanidinium-HCl, 0.1 M Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, 0.01 M Tris-HCl (pH 8.0), 5 mM imidazole, and 10 mM  $\beta$ -mercaptoethanol), and incubated with Ni<sup>2+</sup>-NTA beads (QIAGEN) for 4 h at room temperature. The beads were sequentially washed with buffer A, buffer B (8 M urea, 0.1 M Na<sub>2</sub>PO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, 0.01 M Tris-HCl (pH 8.0), and 10 mM  $\beta$ -mercaptoethanol), and buffer C (8 M urea, 0.1 M Na<sub>2</sub>PO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, 0.01 M Tris-HCl (pH 6.3), and 10 mM  $\beta$ -mercaptoethanol). Bound proteins were eluted with buffer D (200 mM imidazole, 0.15 M Tris-HCl (pH 6.7), 30% glycerol, 0.72 M  $\beta$ -mercaptoethanol, and 5% SDS), and subject to immunoblot analysis. Ubiquitination site prediction software was used for CARM1 ubiquitination site prediction<sup>28</sup>.

**Bacterial expression and GST pull-down assay.** Glutathione S-transferase (GST)-tagged constructs were transformed in Rosetta *Escherichia coli* and purified with glutathione beads (GE Healthcare). <sup>35</sup>S-methionine-labelled TFEB deletions or CARM1 deletions were generated using TNT Quick Coupled Transcription/Translation system (Promega) according to the manufacturer's guidance. Purified proteins and *in vitro* translated proteins were diluted in binding buffer (125 mM NaCl, 20 mM Tris (pH 7.5), 10% glycerol, 0.1% NP-40, 0.5 mM DTT supplemented with protease inhibitors) for GST pull-down experiment. Samples were then washed four times with dilution buffer and boiled with SDS sample buffer for immunoblotting analysis.

**In vitro kinase assay.** GST-SKP2 and beclin (1–148 amino acids) were purified using glutathione bead and eluted in elution buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM L-glutathione reduced (Sigma)). HA-AMPK $\alpha$ 1 constitutively active (CA) was co-transfected in HEK293T cells with Flag-AMPK $\beta$  and HA-AMPK $\gamma$ , and the complex was immunoprecipitated using Flag-M2 beads (Sigma) and eluted through 3 $\times$ -Flag peptide in elution buffer (0.1 mg ml<sup>−1</sup> in TBS). Then 1  $\mu$ g of each substrate was reacted with AMPK complexes in kinase reaction buffer containing 20 mM HEPES (pH 7.4), 5 mM MgCl<sub>2</sub>, 1 mM EGTA, 0.4 mM EDTA and 0.05 mM DTT, as previously described<sup>29</sup>. Reactions were incubated with 150  $\mu$ M AMP and 2  $\mu$ Ci of radiolabelled [ $\gamma$ -<sup>32</sup>P]ATP at 30 °C for 15 min. The reactions were terminated by adding SDS sampling buffer, and phosphorylation was detected by SDS-PAGE and autoradiography.

**Construction of reporter plasmids and luciferase assays.** The *Skp2* promoter region (from 1 kb upstream of transcription start site to 200 bp downstream) and 2 $\times$  CLEAR (GTCAAGTGAACCCAGGGTCAAGTGAC) sequence (underlined bases denote the known sequence of the CLEAR element) were cloned into pGL2-luciferase reporter vector (Promega). FOXO response element (FRE) mutant at the *Skp2* promoter was constructed by site-directed mutagenesis. MEFs were transiently transfected with luciferase reporter plasmids and luciferase activity was measured 36 h after transfection and normalized by  $\beta$ -galactosidase expression.

**qRT-PCR.** Total RNAs were extracted using Trizol (Invitrogen) and reverse transcription was performed from 2.5  $\mu$ g total RNAs using the M-MLV cDNA Synthesis kit (Enzynomics). The abundance of mRNA was detected by an ABI prism 7500 system or BioRad CFX384 with SYBR TOPreal qPCR 2 $\times$  PreMix (Enzynomics). The quantity of mRNA was calculated using the  $\Delta\Delta C_t$  method and *Hprt*, *Gapdh* and *Actb* were used as controls. mRNA levels from mouse liver tissues were normalized by the *36B4* (also known as *Rplp0*) gene. All reactions were performed as triplicates.

The following mouse primers were used in this study. *Actb*; forward (fwd) 5'-TAGCCATCCAGGCTGTGCTG-3', reverse (rev) 5'-CAGGATCTTC ATGAGGTAGTC-3'; *Gapdh*; fwd 5'-CATGGCCTTCCGTGTTCTCA-3', rev 5'-CCTGCTTACCACCTTCTT A-3'; *Hprt*; fwd 5'-GCTGGTGA AAA GGACCTCTCG-3', rev 5'-CCACAGGACTAGAACACCTGC-3'; *36B4*; fwd 5'-CA ACCAGCTCTGGAGAAAC-3', rev 5'-CCAACAGCATATCCCGAATC-3'; *Ulk1*; fwd 5'-GCTCCGGTGACTTACAAAGCTG-3', rev 5'-GCTGACTCCAAG CCAAAGCA-3'; *Map1lc3b*; fwd 5'-CACTGCTCTGCTTGTGTAGGTTG-3',

rev 5'-TCGTTGTGCCTTTATTAGTGCATC-3'; *Atg12*; fwd 5'-TCCGTGCCATCACATACACA-3', rev 5'-TAAGACTGCTGTGGGCTGA-3'; *Atg13*; fwd 5'-CCAGGCTCGACTTGGAGAAAA-3', rev 5'-AGATTTCCACACACATAGATCGC-3'; *Atg14*; fwd 5'-AGCGGTGATTTCTGTCTATTTTCG-3', rev 5'-GCTGTTCAATCCTCATCTTGCAT-3'; *Sirt1*; fwd 5'-GATACCTTGGAGCAGGTTGC-3', rev 5'-CTCCACGAACAGCTTCACAA-3'; *Sqstm1*; fwd 5'-ATGTGGAACATGGAGGGAAGA-3', rev 5'-GGAGTTACACCTGTAGATGGGT-3'; *Vps11*; fwd 5'-AAAAGAGAGACGGTGGCAATC-3', rev 5'-AGCCCAGTAACGGGATAGTTG-3'; *Atp6v1c1*; fwd 5'-ACTGAGTTCTGGCTCATATCTGC-3', rev 5'-TGGAAGAGACGGCAAGATTATG-3'; *Hexb*; fwd 5'-CTGGTGTGCTAGTGTGCG-3', rev 5'-CAGGGCCATGATGTCTCTTGT-3'; *Neu1*; fwd 5'-GGACCGCTGAGCTATTGGG-3', rev 5'-CGGGATGCGGAAAGTGTCTA-3'; *Mcoln1*; fwd 5'-CTGACCCCAATCCTGGGTAT-3', rev 5'-GGCCCGAACTTGTGCAT-3'; *Ctns*; fwd 5'-ATGAGGAGGAATTGGCTGCTT-3', rev 5'-ACGTTGGTTGAACTGCCATTTT-3'; *Hspa5*; fwd 5'-ACTTGGGGACCACCTATTTCCT-3', rev 5'-ATCGCCAATCAGACGCTCC-3'; *Skp2*; fwd 5'-CCTCCAAGGAAACGAGTCAAG-3', rev 5'-CAGGAGACACCTGGAAAGTTC-3'; *Tfeb*; fwd 5'-AAGGTTCCGGAGTATCTGTCTG-3', rev 5'-GGGTTGGAGCTGATATGTAGCA-3'; *Tfe3*; fwd 5'-TGCGTCAGCAGCTTATGAGG-3', rev 5'-AGACAGCCAATCAGAGAT-3'; *Ampk1*; fwd 5'-GTCAAAGCCGACC CAATGA-3', rev 5'-CGTACACGCAAATAATAGGGTT-3'; *Ampk2*; fwd 5'-CAGGCCATAAAGTGGCAGTTA-3', rev 5'-AAAAGTCTGTCGGAGTGCTGA-3'.

The following human primers were used in this study. *ACTB*; fwd 5'-ATGCCGACAGGATGCAGAA-3', rev 5'-ACATCTGCTGGAAGGTGGACAG-3'; *GAPDH*; fwd 5'-CGACCACTTTGTCAAGCTCA-3', rev 5'-AGGGGAGATTACAGTGGGTG-3'; *HPRT*; fwd 5'-TGACACTGGCAAAACAATGCA-3', rev 5'-GGTCTCTTTTACCAGCAAGCT-3'; *SKP2*; fwd 5'-ATGCCCAATCTTGCTCATCT-3', rev 5'-CACCGACTGAGTGATAGGTGT-3'; *AMPK1*; fwd 5'-TTTGCGTGTACGAAGGAAGAAT-3', rev 5'-CTCTGTGGA GTAGCAGTCCCT-3'; *AMPK2*; fwd 5'-CTGTAAGCATGGACGGGTTGA-3', rev 5'-AAATCGGCTATCTTGGCAT TCA-3'.

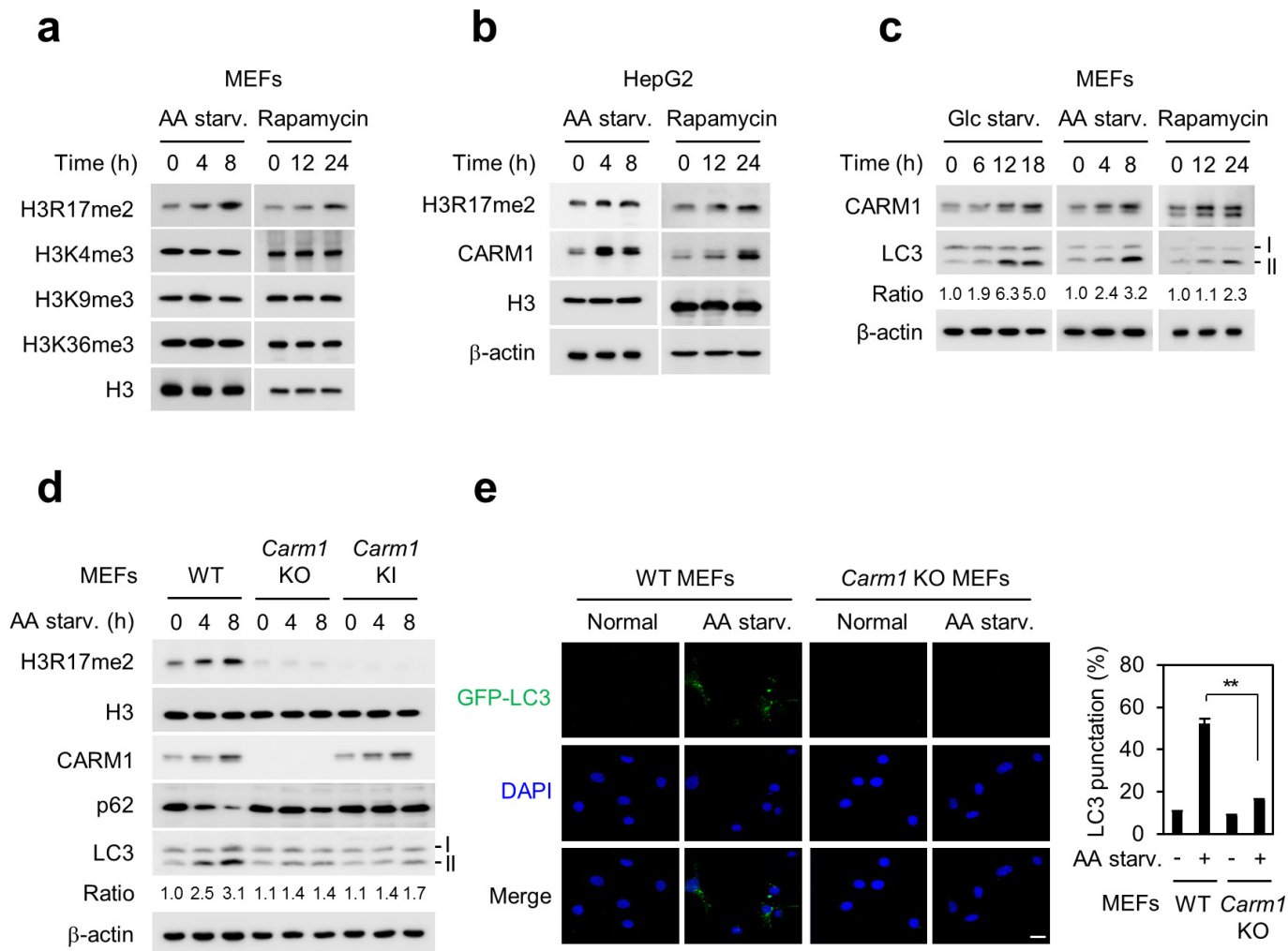
**RNA-seq and ChIP-seq analyses.** The TruSeq method was used to generate RNA-seq libraries. ChIP-seq libraries were prepared using the NEXTFlex ChIP-seq kit (Bioo Scientific), according to the manufacturer's instructions. RNA-seq libraries were pair-end sequenced and ChIP-seq libraries were single-end sequenced on an Illumina Hi-seq 2500 (NICEM, Seoul National University). All the RNA-seq data were mapped using Tophat package<sup>30</sup> against the mouse genome (mm9). Differential analysis has been done via EdgeR package<sup>31</sup>. Differentially regulated genes were identified using a false discovery rate (FDR) cut-off of  $1 \times 10^{-5}$  for knockout against knockout-glucose starvation (KO-GS), wild type against wild-type-glucose starvation (WT-GS), wild type against knockout, and WT-GS against KO-GS. We did hierarchical clustering analysis using the gene expression values from all conditions and replicates for previously selected differential genes. Specifically, we used Ward's criterion for genes with 1 – (correlation coefficient) as a distance measure. Clustering heatmap was drawn using z-score that is scaled across samples for each gene. ChIP-seq data were mapped to the mouse genome using Bowtie. The tracks were generated using uniquely aligned reads. At promoters, genes were sorted based on the expression levels, indicating that H3R17me2 as well as H3K4me3 were enriched at active promoters. We used 8,398 distal (<2.5 kb from annotated TSSs) CBP and MED12 binding sites for enhancers, which were sorted based on H3K27ac levels. H3R17me2 was not detected at enhancers. The data on H3R17me2, H3K4me1, H3K4me3 and H3K27ac were obtained from MEFs under normal conditions.

**ChIP, two-step ChIP assays, and qRT-PCR analyses.** The ChIP and sequential two-step ChIP assays were conducted as previously described<sup>32</sup>. In brief, cells were crosslinked with 1% formaldehyde for 10 min at room temperature. After glycine quenching, the cell pellets were lysed in buffer containing 50 mM Tris-HCl (pH 8.1), 10 mM EDTA, 1% SDS, supplemented with complete protease

inhibitor cocktail (Roche), and sonicated. Chromatin extracts containing DNA fragments with an average of 250 bp were then diluted ten times with dilution buffer containing 1% Triton X-100, 2 mM EDTA, 150 mM NaCl and 20 mM Tris-HCl (pH 8.1) with complete protease inhibitor cocktail, pre-cleared with protein A/G sepharose and subjected to immunoprecipitations overnight at 4 °C. Immunocomplexes were captured by incubating 45 µl of protein A/G sepharose for 2 h at 4 °C. Beads were washed with low-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 150 mM NaCl), high-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 500 mM NaCl), buffer III (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 10 mM Tris-HCl (pH 8.1), 1 mM EDTA), TE buffer (10 mM Tris-HCl (pH 8.0), 0.5M EDTA) and eluted in elution buffer (1% SDS, 0.1 M NaHCO<sub>3</sub>). The supernatant was incubated overnight at 65 °C to reverse-crosslink, and then digested with RNase A for 2 h at 37 °C and proteinase K for 2 h at 55 °C. ChIP and input DNA were then purified and analysed for qRT-PCR analysis or used for constructing sequencing libraries. For the two-step ChIP assays, components were eluted from the first immunoprecipitation reaction by incubation with 10 mM DTT at 37 °C for 30 min and diluted 1:50 in ChIP dilution buffer followed by re-immunoprecipitation with the second antibodies. Two-step ChIP assay was performed in essentially the same way as the first immunoprecipitation. qPCR was used to measure enrichment of bound DNA, and the value of enrichment was calculated relative to input and the ratio to IgG. All reactions were performed in triplicates. The following primers were used in ChIP assays. *Skp2* (FRE); fwd 5'-CCTTAGGACTGGGTCTGTGG-3', rev 5'-GCACGCTGATTGATCTTCA-3'; *Map1lc3b*; fwd 5'-AGCCAGTGGGATATTGGTCT-3', rev 5'-AGAGCTGCGGTACCCTAC-3'; *Atg14*; fwd 5'-GAGACGCCATGATGATCTGA-3', rev 5'-GCCAAGGAGTGTGGGAAGTA-3'; *Atp6v1c1*; fwd 5'-ACTCAGTGGCAGAAGGGAGA-3', rev 5'-AAACACCCAGTGGAGACT GC-3'; *Hexb*; fwd 5'-GAATTGGGACTGTGGTGCAT-3', rev 5'-CTAGTGTGCTGGCCCTA GT-3'; *Hspa5*; fwd 5'-ATTGGTGGCCGTTAAGAATG-3', rev 5'-TGAAGTCGCTACTCGTT GGA-3'; *Ctns*; fwd 5'-CCTCTGGTAGCGTAGGT-3', rev 5'-GCTTTTGGTGAGTCTGTCC-3'; *Vps11*; fwd 5'-GGCCGATCTTAACCTTTGT-3', rev 5'-AGCCAGATGTCTTTTGTGG-3'; *Neu1*; fwd 5'-AGGATGACTCAGCCTGGTG-3', rev 5'-AGGATAGTATGGGCCGAACC-3'; *Mcoln1*; fwd 5'-TACTGGAAGATGGGCTTTCG-3', rev 5'-TGCCAGATTCTAGGAGGAA-3'.

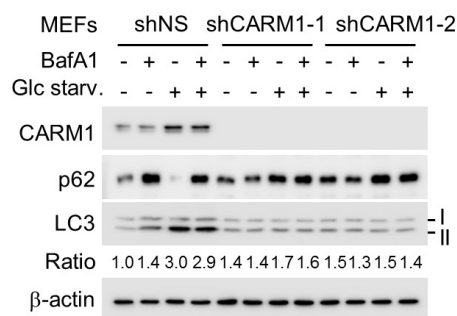
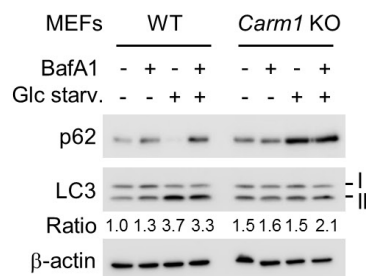
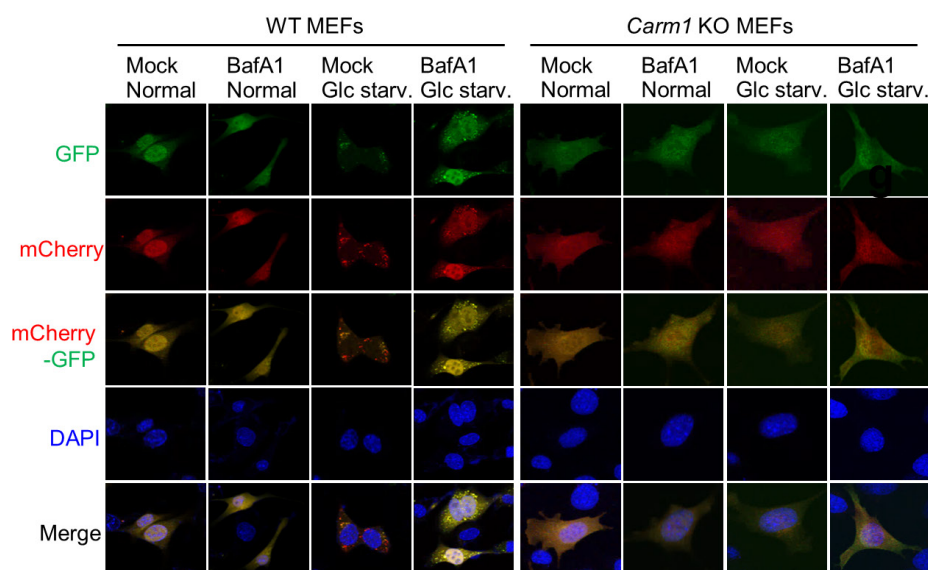
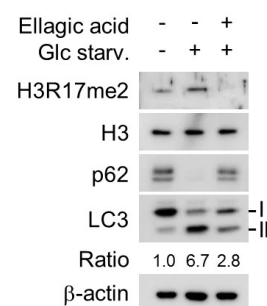
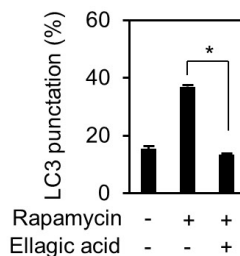
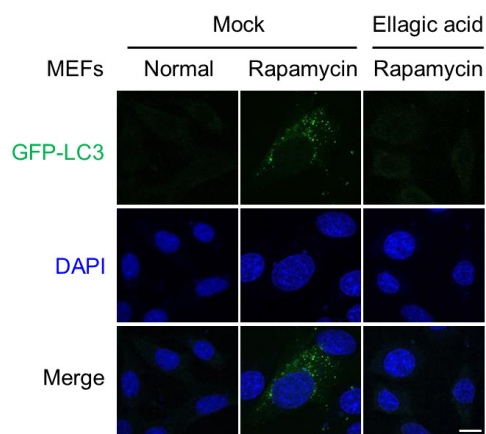
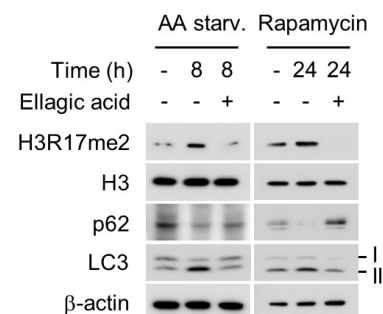
**Statistical analysis.** All experiments were performed independently at least three times. For GFP-LC3 puncta counting, five random confocal images were chosen and the number of cells with GFP-positive dots was counted. An average of 80 cells was examined for each group and *P* values were calculated using one-tailed *t*-tests. For animal studies, sample size for experiments were determined empirically based on previous studies to ensure appropriate statistical power. Mice in the study were randomly chosen for ellagic acid treatment and fasting. No animals were excluded from statistical analysis, and the investigators were not blinded in the studies. Values are expressed as mean ± s.e.m. Significance was analysed using two-tailed, unpaired *t*-test. *P* < 0.05 was considered statistically significant.

26. Kim, I. S. *et al.* Roles of Mis18α in epigenetic regulation of centromeric chromatin and CENP-A loading. *Mol. Cell* **46**, 260–273 (2012).
27. Kim, H. *et al.* DNA damage-induced RORα is crucial for p53 stabilization and increased apoptosis. *Mol. Cell* **44**, 797–810 (2011).
28. Chen, Z., Zhou, Y., Song, J. & Zhang, Z. hCKSAAP\_UbSite: improved prediction of human ubiquitination sites by exploiting amino acid pattern and properties. *Biochim. Biophys. Acta* **1834**, 1461–1467 (2013).
29. Kim, J. *et al.* Differential regulation of distinct Vps34 complexes by AMPK in nutrient stress and autophagy. *Cell* **152**, 290–303 (2013).
30. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
31. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
32. Boo, K. *et al.* Pontin functions as an essential coactivator for Oct4-dependent lincRNA expression in mouse embryonic stem cells. *Nat. Commun.* **6**, 6810 (2015).



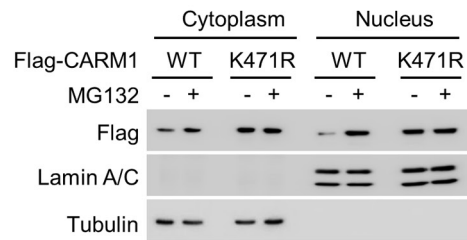
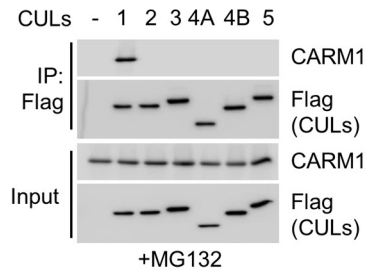
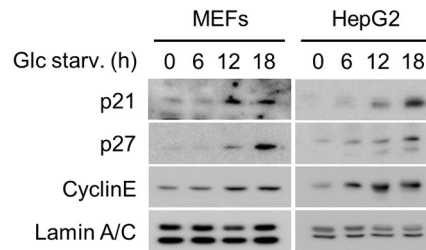
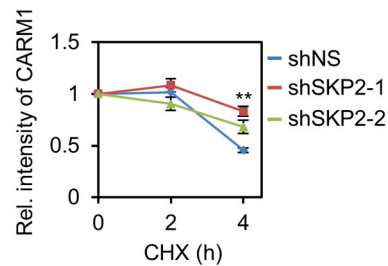
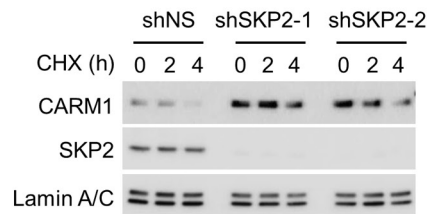
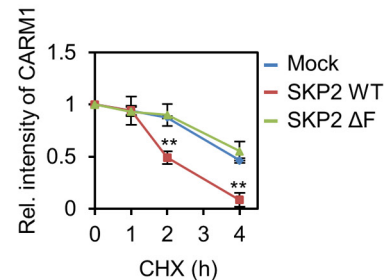
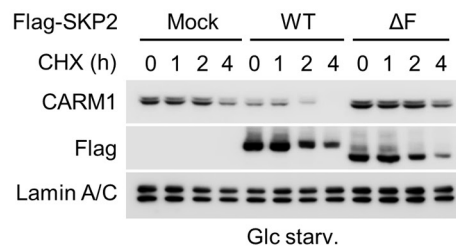
**Extended Data Figure 1 | Increased H3R17me2 by CARM1 in amino acid starvation-induced autophagy.** **a, b,** Immunoblot analysis of various histone marks in response to amino acid (AA) starvation or rapamycin (100 nM). **c,** Immunoblot analysis of CARM1 and LC3 conversion (LC3-II). **d,** Amino acid-starved wild-type, *Carm1* knockout or knock-in

MEFs were analysed by immunoblot. **e,** Representative confocal images of GFP-LC3 puncta formation. GFP-LC3 (green); DAPI (blue). Scale bar, 20  $\mu$ m. The graph shows quantification of LC3-positive punctate cells (right). Bars, mean  $\pm$  s.e.m.;  $n = 5$ , with over 100 cells.  $**P < 0.01$  (one-tailed  $t$ -test).

**a****b****c****d****e****f**

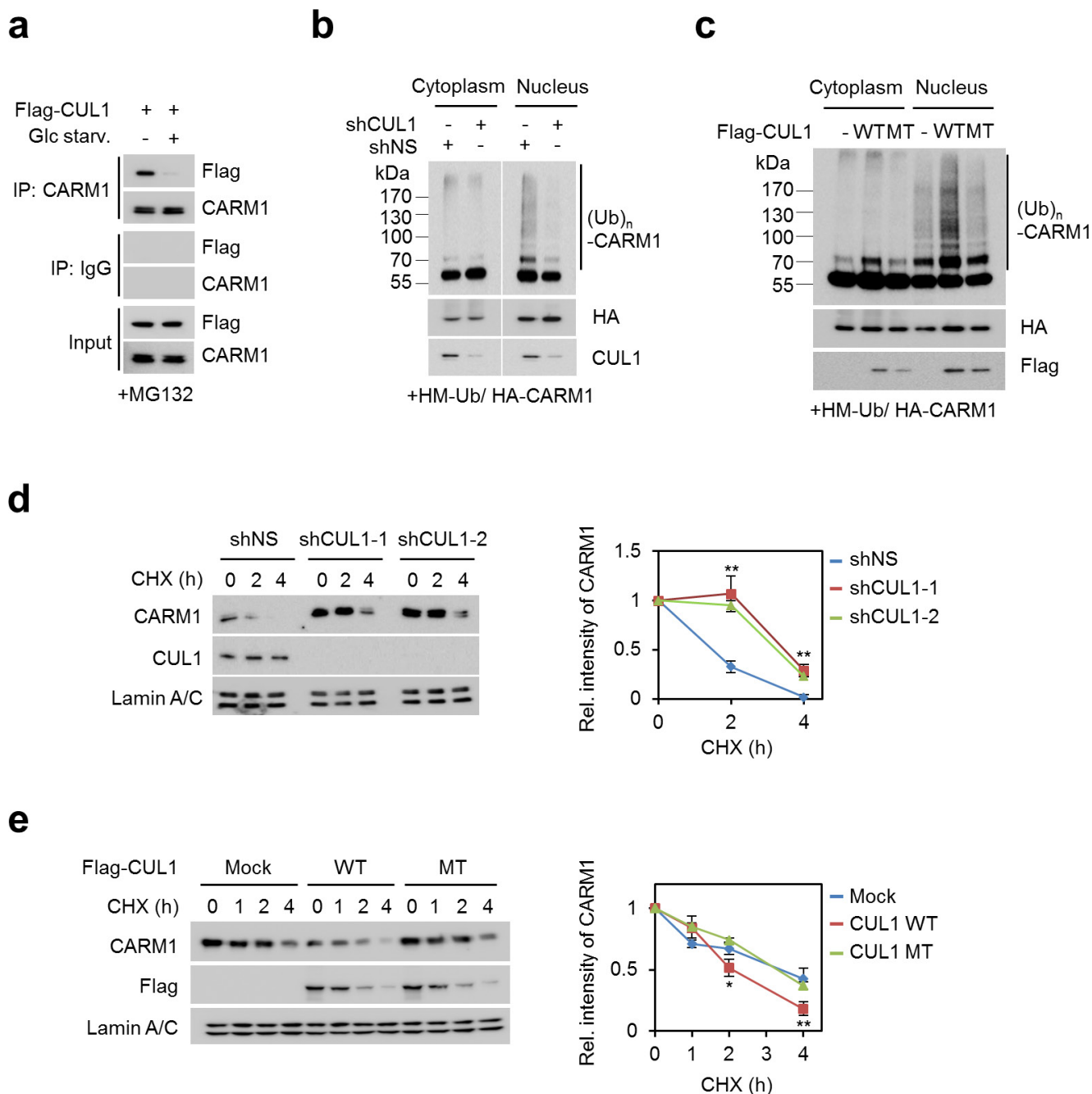
**Extended Data Figure 2 | Loss of CARM1 and inhibition of H3R17me2 impair autophagy.** **a**, LC3 flux was analysed in MEFs infected with nonspecific shRNA (shNS) or CARM1 shRNAs (shCARM1-1 and -2). Bafilomycin A1 (BafA1; 200 nM, 2 h). The LC3-II/LC3-I ratio is indicated. **b**, LC3 flux was analysed in wild-type and *Carm1* knockout MEFs in the absence or presence of Bafilomycin A1. The LC3-II/LC3-I ratio is indicated. **c**, mCherry-GFP-LC3 was transfected in wild-type and *Carm1*

knockout MEFs and the formation of autophagosome (mCherry-positive; GFP-positive) and autolysosome (mCherry-positive; GFP-negative) was examined. Scale bar, 20  $\mu$ m. **d**, Immunoblot analysis in MEFs. **e**, Representative confocal images of GFP-LC3 puncta formation. Scale bar, 10  $\mu$ m. Bars, mean  $\pm$  s.e.m.;  $n = 5$ , over 150 cells. \* $P < 0.05$  (one-tailed  $t$ -test). **f**, Immunoblot analysis in MEFs.

**a****b****c****d****e**

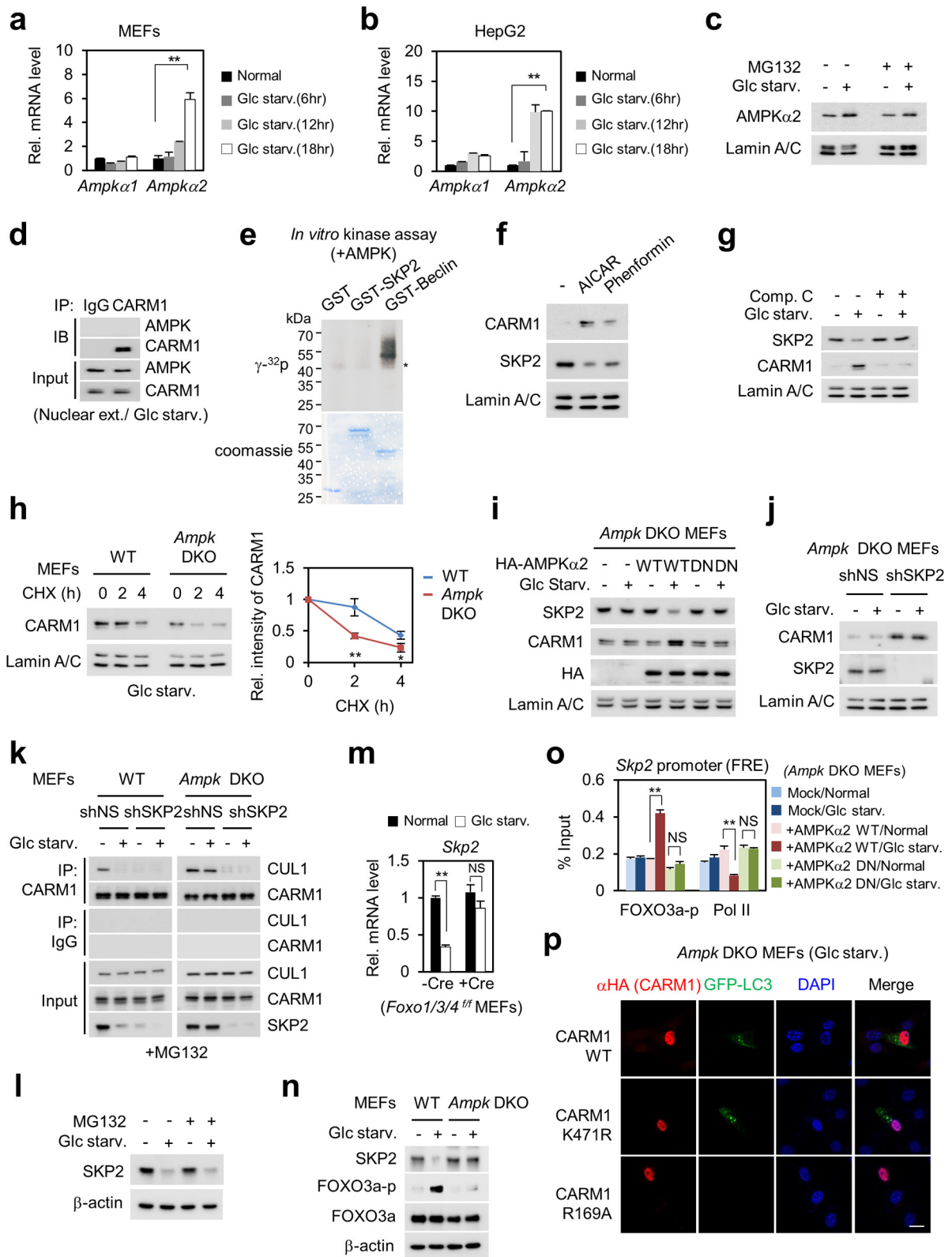
**Extended Data Figure 3 | CARM1 is degraded by SKP2-containing SCF E3 ligase in the nucleus.** **a**, Wild-type CARM1 and ubiquitination-defective mutant K471R were analysed for their expression in MEFs after MG132 treatment. **b**, Interaction between CARM1 and CUL proteins was analysed. **c**, Lysates were analysed by immunoblot. **d**, Left, HepG2 cells

infected with two different SKP2 shRNAs were subject to cycloheximide (CHX) experiment. Right, protein half-life of CARM1 was quantitatively defined (right). **e**, Left, CHX experiment in HepG2 expressing wild-type SKP2 or ΔF mutant. Right, protein half-life of CARM1 was quantitatively defined. Data are mean  $\pm$  s.e.m.;  $n = 3$ . \*\* $P < 0.01$  (one-tailed  $t$ -test) (**d**, **e**).



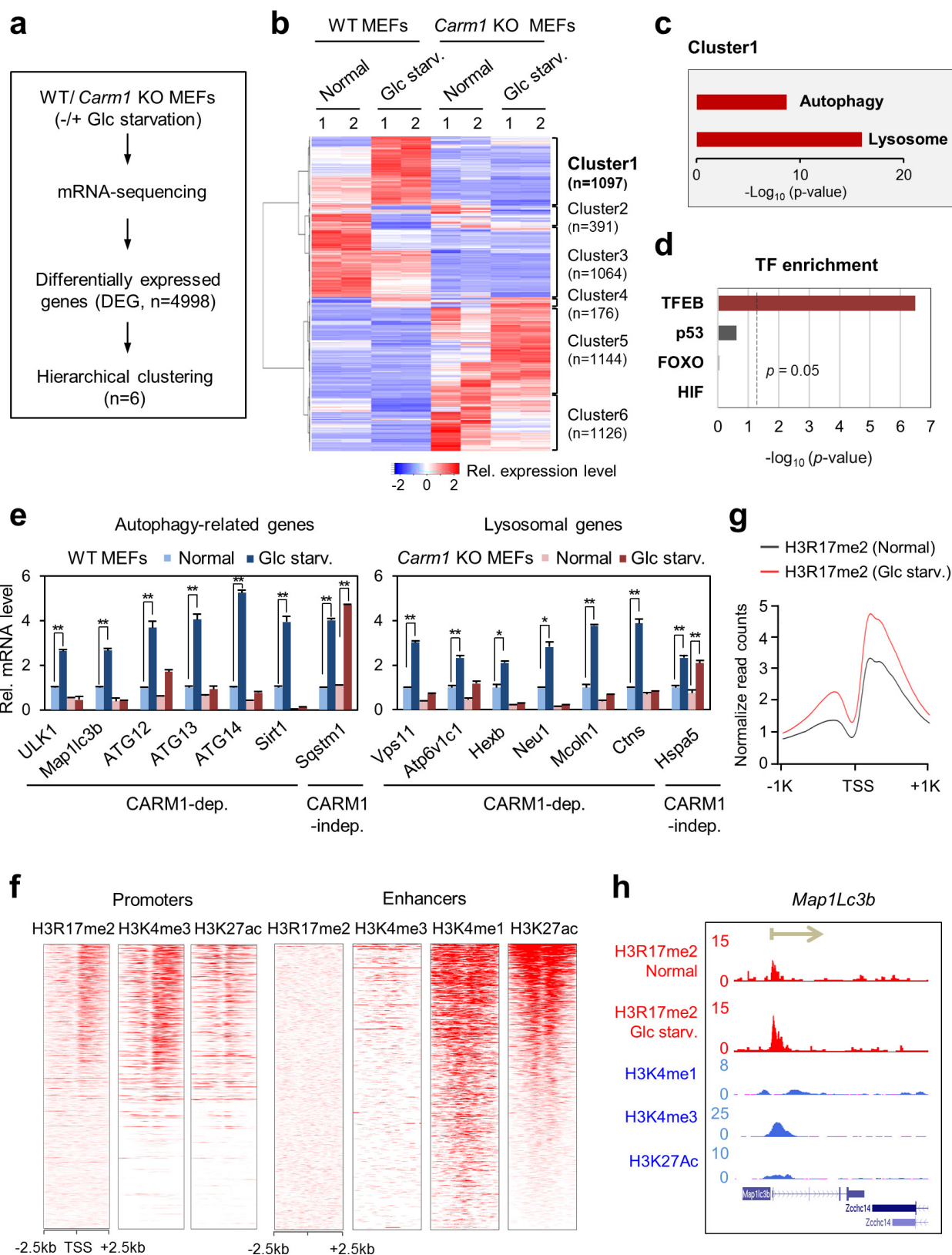
**Extended Data Figure 4 | CARM1 is degraded by CUL1-containing SCF E3 ligase in the nucleus under nutrient-rich condition.** **a**, HepG2 cells transfected with Flag-CUL1 were deprived of glucose for 18 h and treated with MG132 before collecting. Interaction between CARM1 and CUL1 was analysed. **b, c**, *In vivo* ubiquitination assay of CARM1 after knockdown of CUL1 (**b**) or overexpression of wild-type or K720R

mutant (MT) CUL1 (**c**). **d, e**, Left, HepG2 cells infected with two different CUL1 shRNAs (**d**) or overexpressing wild-type or mutant CUL1 (**e**) were subject to cycloheximide treatment. Right, protein half-life of CARM1 was quantitatively defined. Data are mean  $\pm$  s.e.m.;  $n = 3$ . \* $P < 0.05$ , \*\* $P < 0.01$  (one-tailed *t*-test) (**d, e**).



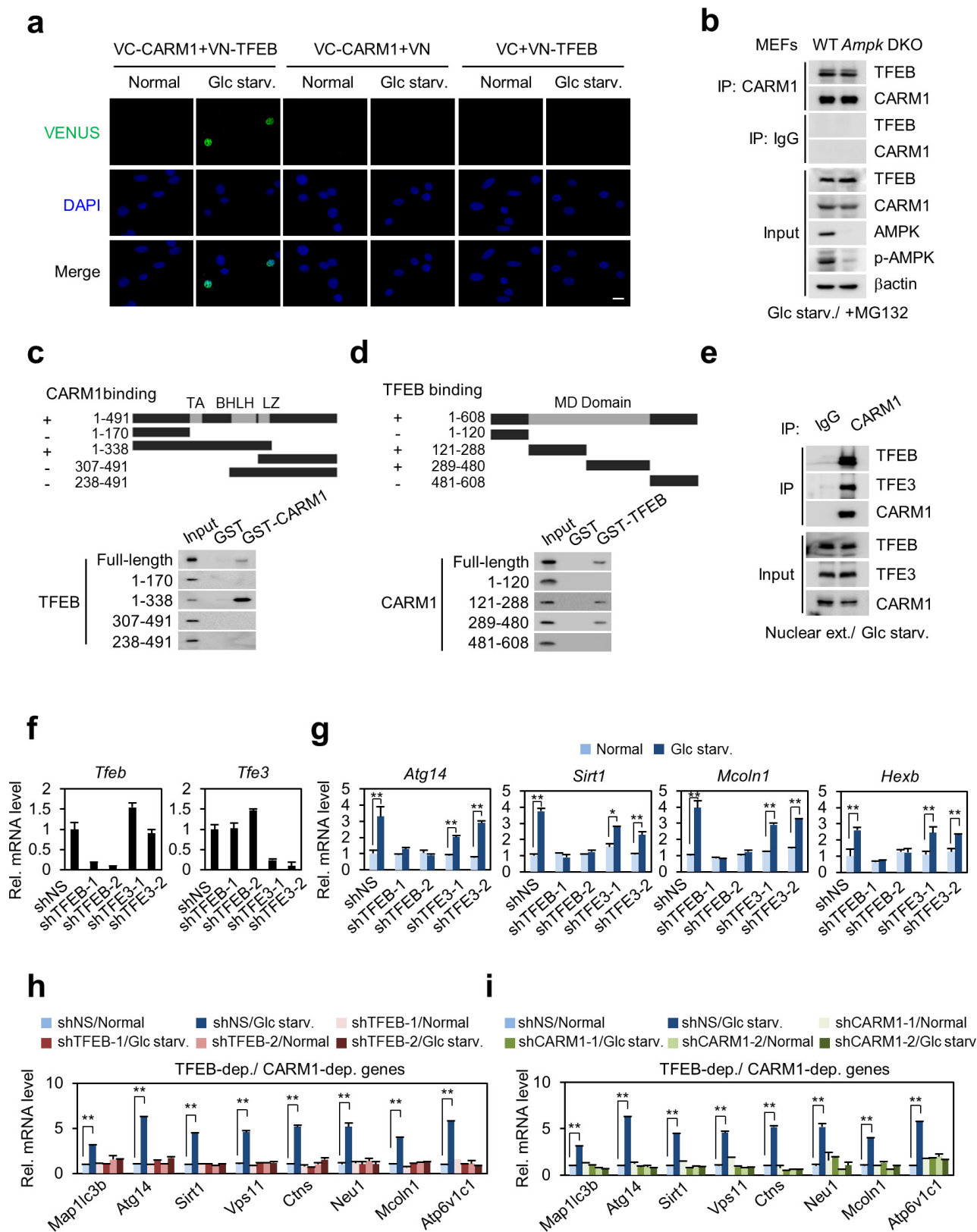
**Extended Data Figure 5 | AMPKα2 accumulates in the nucleus leading to repression of SKP2 and stabilization of CARM1 under nutrient-starved conditions.** **a, b**, qRT-PCR of *Ampkα1* and *Ampkα2* in MEFs (**a**) and HepG2 cells (**b**) upon glucose starvation. **c**, The nuclear AMPKα2 expression level was analysed in the absence or presence of MG132. **d**, Binding between CARM1 and AMPK was assessed. **e**, In vitro kinase assay with AMPK. **f**, MEFs were treated with AICAR (1 mM) or phenformin (2 mM) for 4 h. The nuclear fraction was analysed by immunoblot. **g**, MEFs were deprived of glucose in the absence or presence of 10 μM compound C and the nuclear fraction was analysed

by immunoblot. **h**, Left, cycloheximide treatment in wild-type and *Ampk* DKO MEFs. Right, protein half-life of CARM1 was quantitatively defined. **i, j**, *Ampk* DKO MEF lysates were analysed by immunoblot. **k**, CARM1–CUL1 interaction was analysed after SKP2 knockdown in wild-type and *Ampk* DKO MEFs. **l**, SKP2 expression levels were analysed in the absence or presence of MG132. **m**, *Foxo1/3/4*<sup>+/+</sup> MEFs infected with Cre virus were analysed for *Skp2* mRNA. **n**, SKP2 and phosphorylated FOXO3a were analysed by immunoblot. **o**, ChIP assay of the *Skp2* promoter. Data are mean ± s.e.m.; *n* = 3. \**P* < 0.05, \*\**P* < 0.01 (one-tailed *t*-test) (**a, b, h, m, o**). **p**, Representative confocal images. Scale bar, 20 μm.



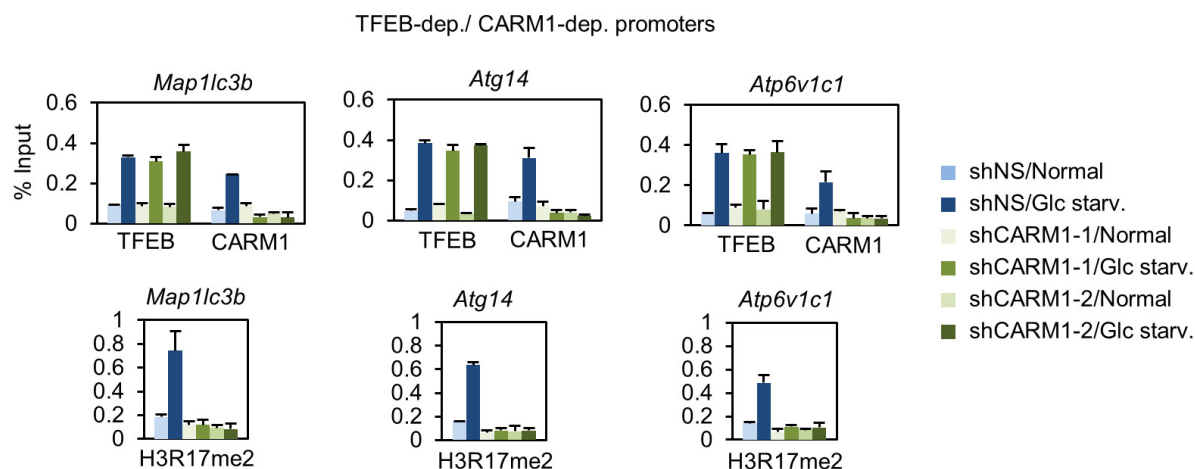
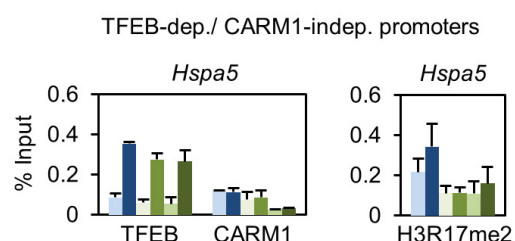
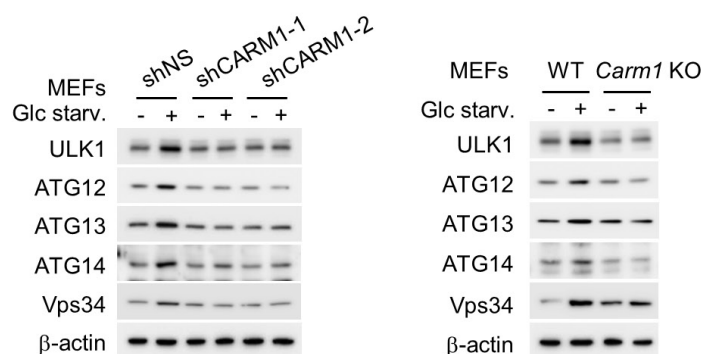
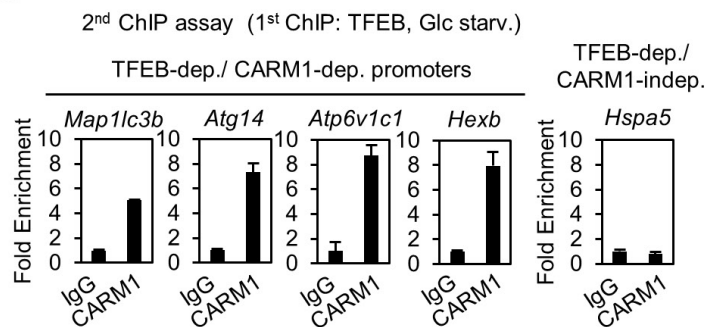
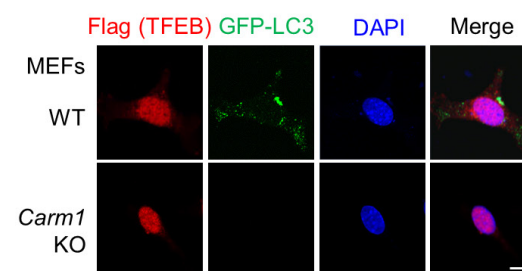
**Extended Data Figure 6 | Identification of CARM1 target genes by RNA-seq and ChIP-seq analyses.** **a**, Flow chart showing the strategy of RNA-seq analysis. **b**, Hierarchical clustering results applied to 4,998 differentially expressed genes (DEGs). **c**, Autophagy-related and lysosomal genes significantly observed in cluster 1. Hyper-geometric  $P$  values were calculated. **d**, Genes from cluster 1 were analysed for transcription factor (TF) motif enrichment at their promoter region ( $-500$ – $100$ ). Hypergeometric  $P$  values were calculated. **e**, qRT-PCR analysis of CARM1-dependent autophagy-related and lysosomal

genes. Data are mean  $\pm$  s.e.m.;  $n = 3$ . \* $P < 0.05$ , \*\* $P < 0.01$  (one-tailed  $t$ -test). **f**, Enrichment of H3R17me2 at promoters (left) and enhancers (right). The data on H3R17me2, H3K4me1, H3K4me3 and H3K27ac were obtained from MEFs under normal condition. **g**, Increase in H3R17me2 at promoters of genes from cluster 1 after glucose starvation. **h**, Increased H3R17me2 levels in response to 18 h of glucose starvation at the autophagy-related gene *Map1lc3b*. The direction of transcription is indicated by the arrow and the beginning of the arrow indicates the TSS.



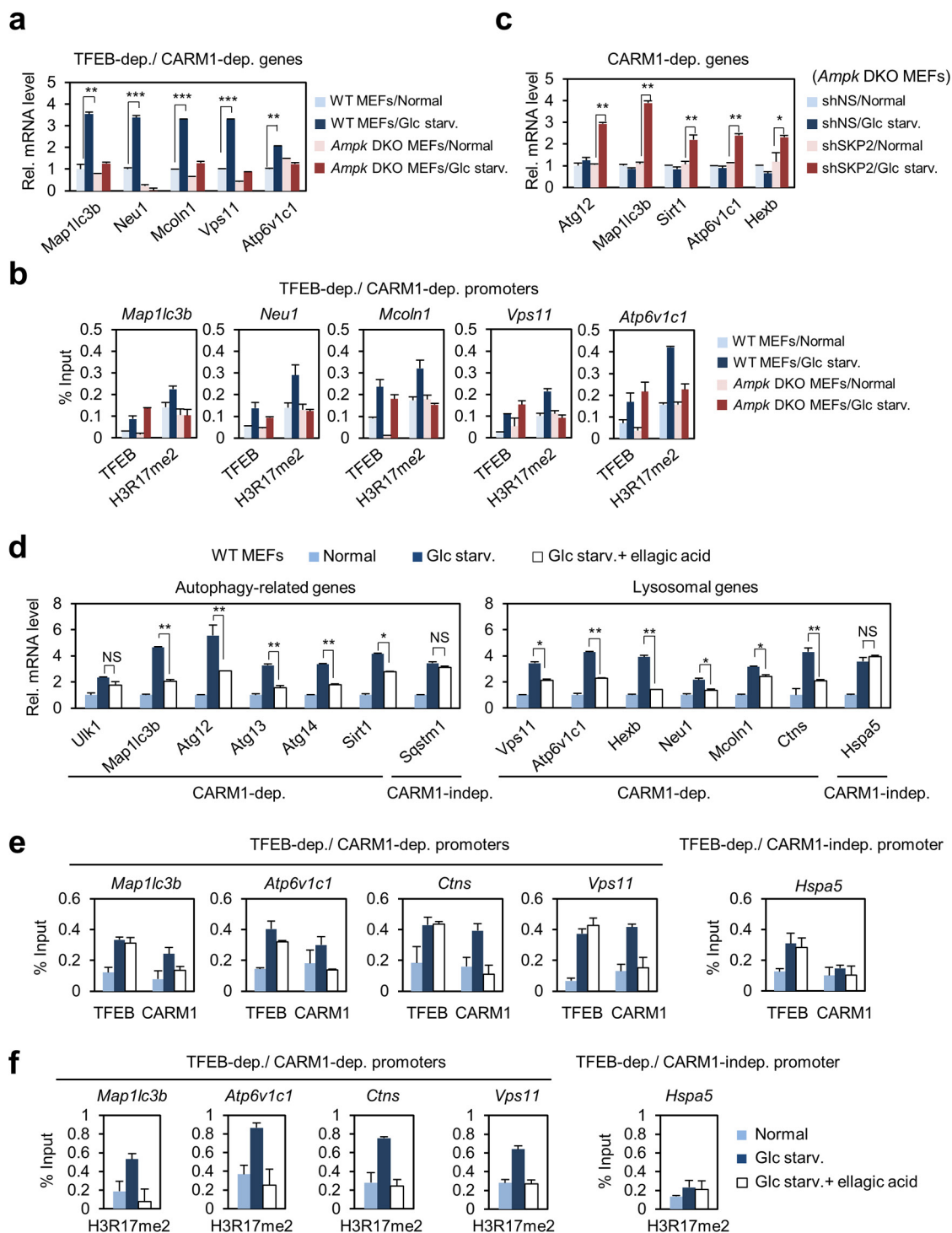
**Extended Data Figure 7 | Binding mapping of CARM1 and TFEB and their target gene regulation in glucose starvation.** **a**, Bimolecular fluorescence complementation (BiFC) analysis of the CARM1–TFEB interaction. Scale bar, 20  $\mu$ m. **b**, Interaction between CARM1 and TFEB was analysed in wild-type and *Ampk* DKO MEFs after glucose starvation. **c**, **d**, *In vitro* GST pull-down assays for domain mapping of CARM1–TFEB interaction. BHLH, basic helix–loop–helix; LZ: leucine zipper.

MD, methyltransferase domain; TA, transcription activation domain. **e**, Endogenous co-immunoprecipitation from nuclear fraction of wild-type MEFs. **f**, **g**, qRT–PCR analysis in MEFs after knockdown of TFEB or TFE3. **h**, **i**, qRT–PCR analysis showing mRNA levels of TFEB-dependent and CARM1-dependent genes after knockdown of TFEB (**h**) or CARM1 (**i**). Bars, mean  $\pm$  s.e.m.;  $n = 3$ . \* $P < 0.05$ , \*\* $P < 0.01$  (one-tailed *t*-test) (**f**–**i**).

**a****b****c****d****e****Extended Data Figure 8 | CARM1 functions as a co-activator of TFEB.**

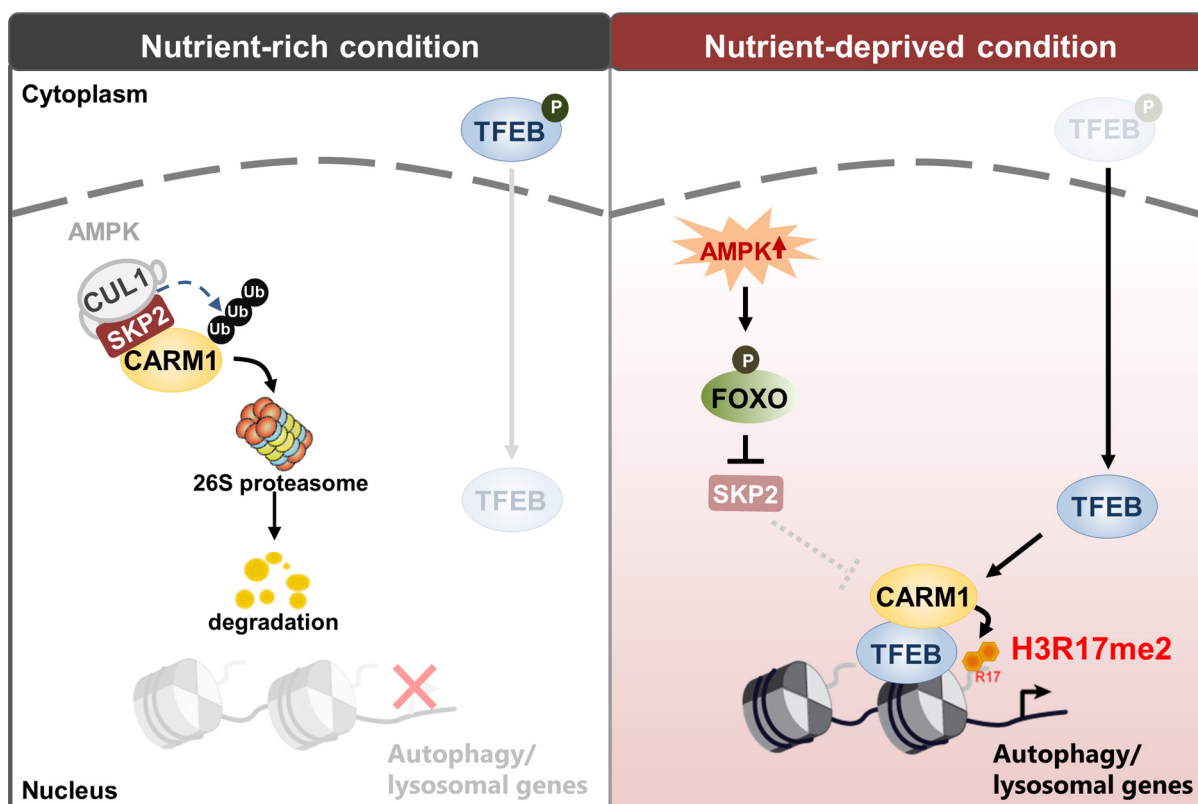
**a**, ChIP assays on TFEB-dependent, CARM1-dependent promoters after knockdown of CARM1. **b**, ChIP assays of the *Hspa5* promoter, a TFEB-dependent, CARM1-independent target promoter. **c**, MEFs were analysed with indicated antibodies. **d**, Two-step ChIP assays were performed on promoters of TFEB-dependent, CARM1-dependent target

genes or TFEB-dependent, CARM1-independent target genes in MEFs after 18 h of glucose starvation. The chromatin fractions were first subject to pull-down with anti-TFEB antibody, eluted from immunocomplexes and applied for the second pull-down with control IgG or anti-CARM1 antibody. Bars, mean  $\pm$  s.e.m.;  $n = 3$  (**a**, **b**, **d**). **e**, Representative confocal images. Scale bar, 10  $\mu$ m.



**Extended Data Figure 9 | A subset of autophagy-related and lysosomal genes regulated by TFEB requires CARM1.** **a**, qRT-PCR analysis showing mRNA levels of TFEB-dependent and CARM1-dependent autophagy-related and lysosomal genes in wild-type and *Ampk* DKO MEFs in response to glucose starvation. **b**, ChIP assays on TFEB-dependent, CARM1-dependent target genes in wild-type and *Ampk* DKO MEFs. **c**, qRT-PCR analysis of CARM1-dependent genes after knockdown of

SKP2 in *Ampk* DKO MEFs. **d**, qRT-PCR analysis was performed in MEFs deprived of glucose in the absence or presence of H3R17me2-specific inhibitor, ellagic acid. **e**, **f**, ChIP assays on TFEB-dependent, CARM1-dependent promoters. *Hspa5* promoter was also analysed as a CARM1-independent promoter. Bars, mean  $\pm$  s.e.m.;  $n = 3$ . \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (one-tailed *t*-test) (**a–f**).



**Extended Data Figure 10 | Graphical summary of the AMPK-SKP2-CARM1 signalling cascade.** Proposed model depicting the AMPK-SKP2-CARM1 signalling axis in the transcriptional and epigenetic regulation of autophagy. The SKP2-containing SCF E3 ubiquitin ligase complex

degrades CARM1 under nutrient-rich conditions, but in nutrient-deprived conditions, AMPK-dependent phosphorylation of FOXO3a downregulates SKP2 and stabilizes CARM1, which in turn functions as a co-activator of TFEB in regulation of autophagy.

# Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor

Shintaro Iwasaki<sup>1</sup>, Stephen N. Floor<sup>1</sup> & Nicholas T. Ingolia<sup>1</sup>

Rocaglamide A (RocA) typifies a class of protein synthesis inhibitors that selectively kill aneuploid tumour cells and repress translation of specific messenger RNAs<sup>1–4</sup>. RocA targets eukaryotic initiation factor 4A (eIF4A), an ATP-dependent DEAD-box RNA helicase; its messenger RNA selectivity is proposed to reflect highly structured 5' untranslated regions that depend strongly on eIF4A-mediated unwinding<sup>5</sup>. However, rocaglate treatment may not phenocopy the loss of eIF4A activity, as these drugs actually increase the affinity between eIF4A and RNA<sup>1,2,6</sup>. Here we show that secondary structure in 5' untranslated regions is only a minor determinant for RocA selectivity and that RocA does not repress translation by reducing eIF4A availability. Rather, *in vitro* and in cells, RocA specifically clamps eIF4A onto polypurine sequences in an ATP-independent manner. This artificially clamped eIF4A blocks 43S scanning, leading to premature, upstream translation initiation and reducing protein expression from transcripts bearing the RocA–eIF4A target sequence. In elucidating the mechanism of selective translation repression by this lead anti-cancer compound, we provide an example of a drug stabilizing sequence-selective RNA–protein interactions.

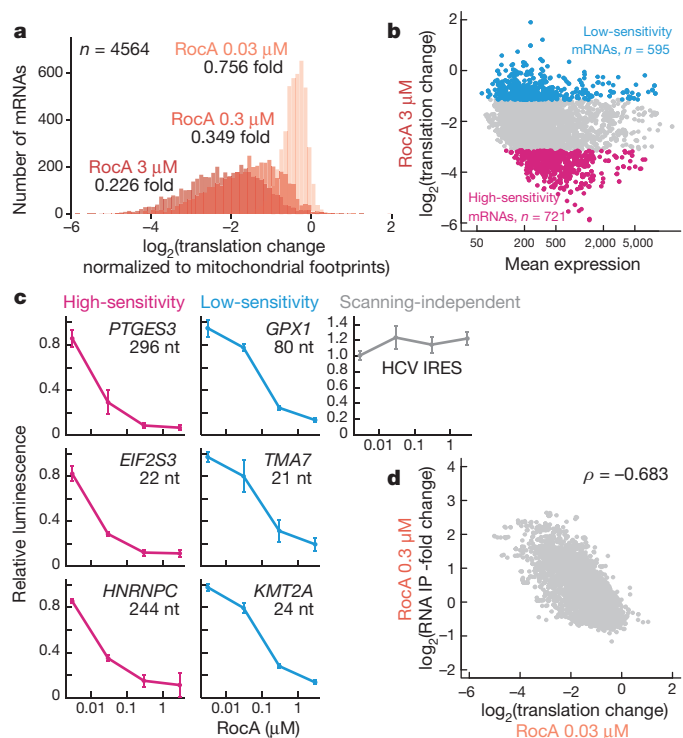
We analysed the global translational inhibition caused by RocA, as well as its marked messenger RNA (mRNA) selectivity, using ribosome profiling<sup>7</sup>. RocA treatment of HEK 293 cells caused a dose-dependent decrease in polysome formation and protein synthesis (Extended Data Figs 1a and 2a). Translation was inhibited without 4EBP dephosphorylation or eIF2 $\alpha$  phosphorylation (Extended Data Fig. 1b), but partly rescued by expression of RocA-resistant eIF4A proteins<sup>6</sup> (Extended Data Fig. 1c, d). We quantified the reduction in overall cytosolic ribosome footprints after normalization of our ribosome profiling data against footprints from the mitochondrial ribosome<sup>8</sup>, which employs molecular machinery distinct from the cytoplasmic translation apparatus (Fig. 1a and Extended Data Fig. 1e–h). We saw that RocA sensitivity varied widely across different transcripts (Fig. 1a, b and Supplementary Table 1a, b). This mRNA-specific translational repression occurred even at a low, therapeutically relevant concentration of RocA (30 nM)<sup>1–4</sup>, correlated well between different drug concentrations, and was not accompanied by significant changes in mRNA abundance (Extended Data Fig. 2b–d and Supplementary Table 1c).

Given that eIF4A acts during the scanning of the pre-initiation 43S complex along the 5' untranslated region (UTR)<sup>9</sup>, we reasoned that the varied RocA sensitivity of different mRNAs might be determined by their 5' UTR sequences. We confirmed that the 5' UTRs of selected mRNAs were sufficient to confer RocA sensitivity on a *Renilla* luciferase reporter, while the scanning-independent HCV IRES<sup>10</sup> was totally resistant to the drug (Fig. 1c and Extended Data Fig. 2e). However, RocA sensitivity did not reflect either the calculated thermodynamic stability or experimentally derived DMS-Seq secondary structure measurement<sup>11</sup> of the 5' UTR, and the presence of predicted G-quadruplexes<sup>5</sup> contributed only modestly (Extended Data Fig. 3).

Because RocA enhances the RNA affinity of eIF4A<sup>1,2,6</sup>, we suspected that it could induce effects beyond the simple loss of eIF4A activity.

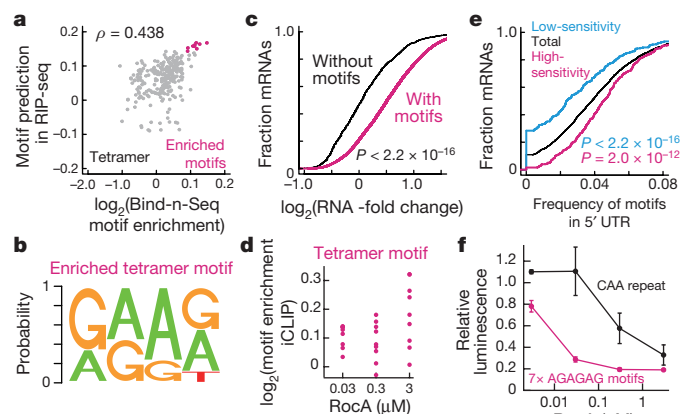
Indeed, we found that the eIF4A inhibitor hippuristanol (Hipp), which decreases the affinity between eIF4A and RNA<sup>12,13</sup>, yields a different spectrum of mRNA-specific repression (Extended Data Fig. 4a–e). The mTOR inhibitor PP242, which inhibits formation of eIF4F (a complex of eIF4E/G/A)<sup>14,15</sup>, represses a subset of these Hipp-sensitive mRNAs (Extended Data Fig. 4f, g). Thus, RocA exerts effects beyond reduced eIF4A activity, particularly at low, therapeutic doses.

We next asked how RocA affected eIF4A occupancy across the transcriptome in cells by sequencing transcripts that co-purified with streptavidin binding peptide (SBP)-tagged eIF4A (Extended Data Fig. 5) (RNA-immunoprecipitation sequencing (RIP-seq)). Increasing



**Figure 1 | RNA sequence selectivity is imparted upon eIF4A by RocA causing selective translation repression.** **a**, Histogram of the number of transcripts along translation -fold change by ribosome profiling when cells are treated with 0.03, 0.3, or 3  $\mu$ M RocA, normalized to the number of mitochondrial footprints. Median -fold change is shown. Bin width is 0.1. **b**, MA plot of mean footprint reads between 3  $\mu$ M RocA treatment and non-treatment normalized to library sizes versus translation -fold change by 3  $\mu$ M RocA treatment, highlighting high- and low-sensitivity mRNAs. **c**, The 5' UTRs of indicated genes were fused to *Renilla* luciferase and these reporter mRNAs were transfected before treatment with RocA as indicated. Data represent mean and s.d. ( $n = 3$ ). **d**, Correlation of translation -fold change to RIP -fold change with RocA treatment.  $\rho$ , Spearman's rank correlation.

<sup>1</sup>Department of Molecular and Cell Biology, Center for RNA Systems Biology, University of California, Berkeley, California 94720, USA.



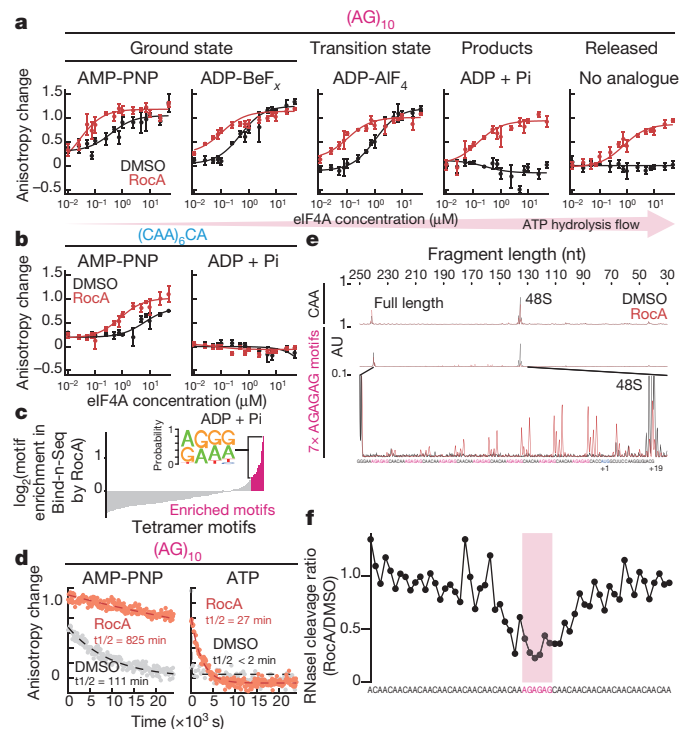
**Figure 2 | RNA Bind-n-Seq and iCLIP reveal that RocA preferentially increases the affinity between eIF4A and polypurine motif.**

**a**, Correlations between tetramer motif enrichment in Bind-n-Seq by 0.03 μM RocA treatment and motif prediction of 0.03 μM RocA effect in RIP-seq. **b**, Highest-scoring elements in Bind-n-Seq and RIP-seq. **c**, The change in mRNA binding for mRNAs with or without the enriched tetramer motif (**b**) in their 5' UTRs is shown as the RIP -fold change by RocA normalized to spike-in RNA. Significance is calculated by Mann–Whitney *U*-test. **d**, Enrichment of tetramer motifs (**b**) in iCLIP by RocA treatment relative to control dimethylsulfoxide (DMSO) treatment. **e**, The frequency of the tetramer motif (**b**) in the 5' UTR predicts whether a mRNA is high- or low-sensitivity, on the basis of the difference in cumulative distributions of motifs in the 5' UTR. Significance is calculated by Mann–Whitney *U*-test. **f**, Reporter assay in HEK 293 cells with a CAA-repeat 5' UTR containing seven polypurine motif (AGAGAG) insertions (Extended Data Fig. 9a). Data represent mean and s.d. (*n* = 3).

RocA doses elevated the overall amount of RNA that co-purified with SBP-tagged eIF4A (Extended Data Fig. 5d), and greatly changed the abundance of individual transcripts, leading to 15-fold or larger differences between mRNAs. Strikingly, enhanced eIF4A binding in the presence of RocA correlated strongly with translation inhibition by RocA (Fig. 1d and Extended Data Fig. 5f), suggesting that a selective increase of the eIF4A–RNA affinity underlies the specific translation inhibition caused by RocA.

This mRNA selectivity led us to explore the sequence preferences of eIF4A in the absence and presence of RocA. We measured the RNAs that bound to eIF4A out of a random pool of oligonucleotides using deep sequencing (RNA Bind-n-Seq)<sup>16</sup> (Extended Data Fig. 6a–c). We then calculated the enrichment of 4- to 6-nucleotide (nt) motifs in RNAs retained on eIF4A, as DEAD-box RNA helicases typically contact 6 nt (ref. 17). The motifs enriched from randomized synthetic RNA by Bind-n-Seq also predicted RIP-seq enrichments of endogenous transcripts (Fig. 2a and Extended Data Fig. 6d). In both experiments, RocA greatly enhanced binding to short polypurine sequences (Fig. 2b, c and Extended Data Fig. 6e). Although drug-free eIF4A also had intrinsic RNA sequence preferences<sup>18</sup> (Extended Data Figs 6g) and transcripts containing these preferred sequences were relatively resistant to Hipp treatment (Extended Data Fig. 6h), RocA only selectively increases binding to a subset of sequences containing polypurine stretches (Extended Data Fig. 6g).

Polypurine motifs were also enriched in the eIF4A binding sites detected by photocrosslinking and immunoprecipitation (iCLIP)<sup>19</sup> after RocA treatment (Fig. 2d and Extended Data Fig. 7), and in the 5' UTRs of translationally RocA-sensitive mRNAs (Fig. 2e). This striking correspondence among *in vitro* binding to recombinant protein, *ex vivo* co-purification, crosslinking in cells, and translational repression in cells led us to hypothesize that selective binding to polypurine motifs induced by RocA binding could explain mRNA-specific translational repression. We then directly confirmed that inserting the polypurine motif into an unstructured CAA repeat 5' UTR (Extended Data Fig. 9a)<sup>20</sup> sensitized the reporter to RocA inhibition (Fig. 2f).

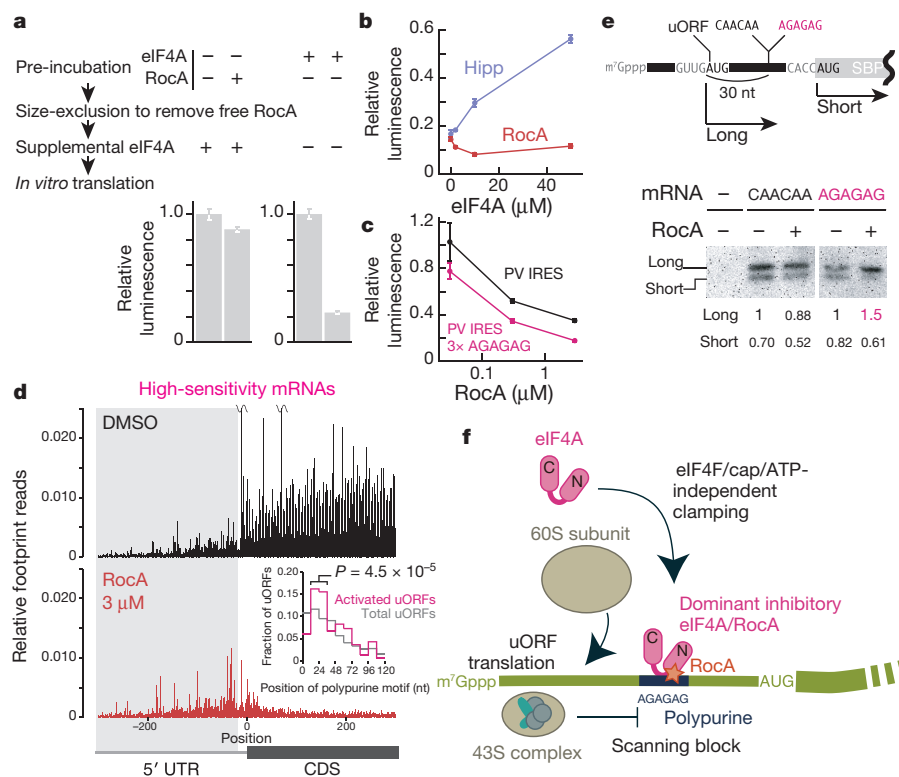


**Figure 3 | RocA clamps eIF4A on polypurine motif even after ATP hydrolysis.**

**a**, **b**, Direct measurement of the eIF4A/RNA affinity by fluorescence polarization for eIF4A and 5' FAM-labelled RNAs in the presence or absence of RocA. Data represent mean and s.d. (*n* = 3). **c**, Motif enrichments along entire tetramer motifs in Bind-n-Seq with ADP + Pi and highest-scoring elements (inset). **d**, Competition assay with unlabelled RNA. Data represent mean (*n* = 3). **e**, Ribosome toeprinting assay performed in RRL in the presence of GMP-PNP in the presence or absence of 3 μM RocA treatment. **f**, Relative RNase I cleavage protected by eIF4A/RocA complex on mRNA containing one AGAGAG at the middle in footprinting assay. See the original data in Extended Data Fig. 9g.

We found that RocA-induced, sequence-selective eIF4A binding occurs through ATP-independent clamping that suffices to repress translation of the clamped mRNA. The cycle of ATP-dependent RNA binding and subsequent release upon ATP hydrolysis is necessary for the efficient RNA remodelling activity of eIF4A as well as its role in translation<sup>21</sup>. Drug-free eIF4A bound RNA only in the presence of ATP (AMP-PNP and ADP-BeF<sub>3</sub>) and transition state (ADP-AlF<sub>4</sub>) analogues but not hydrolysis products (ADP + Pi). Remarkably, RocA clamped eIF4A on polypurine RNA, but not CAA-repeat RNA, in an ATP-independent manner (Fig. 3a, b and Extended Data Fig. 8a, b). Bind-n-Seq performed with ADP + Pi likewise recovered polypurine-enriched RNAs in the presence of RocA and no detectable RNA in the absence of RocA (Fig. 3c and Extended Data Fig. 6i). RocA provided polypurine-specific RNA binding activity to mutant eIF4A defective for ATP binding (VX<sub>4</sub>GKT)<sup>22</sup>, which does not bind to RNA at all without RocA (Extended Data Fig. 8d–f), and even to the truncated amino (N)-terminal domain of eIF4A, albeit with lower affinity (Extended Data Fig. 8g). The eIF4A/RocA complex dissociated far more slowly from polypurine RNA than naive eIF4A, even in the presence of ATP, whose hydrolysis ordinarily permits rapid dissociation (Fig. 3d). High RNA affinity in the ADP-bound state can prolong RNA binding beyond the time required for adenosine nucleotide exchange to restore the high-affinity ATP-bound state and thus greatly reduce the effective dissociation rate. This effective dissociation rate from polypurine RNA measured in hydrolysable ATP (reflective of the intracellular environment) becomes much slower than the ~1 min timescale of translation initiation<sup>23</sup>, and could serve to directly block the ribosome.

To probe how clamped eIF4A repressed translation, we recapitulated RocA-induced, polypurine motif-specific translational repression



**Figure 4 | eIF4A/RocA complexes on polypurine motifs block scanning of pre-initiation complex, inducing uORF translation.** **a**, Pre-formation of the complex with RocA and eIF4A on the mRNA bearing seven polypurine motifs represses the translation from the mRNA in RRL. **b**, The supplementation of recombinant eIF4A protein to RRL *in vitro* translation reaction with 10 μM Hipp or 3 μM RocA. **c**, *In vitro* translation in RRL with mRNAs with native polio virus IRES and that with three polypurine motifs (Extended Data Fig. 9a). **d**, Meta-gene analysis of high-sensitivity transcripts to RocA. Reads are normalized to the sum of mitochondrial footprints reads. Histogram of the position of the first polypurine motif (hexamer) after uORF initiation codon (inset). *P* value is

in rabbit reticulocyte lysate (RRL) (Extended Data Fig. 9b, c). In this system, RocA treatment represses the formation of 48S pre-initiation complexes on the start codon of sensitive mRNAs, which we assessed using a primer extension toeprinting assay<sup>20,24</sup> (Fig. 3e and Extended Data Fig. 9d). Surprisingly, we observed additional RocA-dependent toeprints on the 5' UTR, corresponding to the position of polypurine motifs (Fig. 3e), even without eIF4F recruitment (Extended Data Fig. 9e). We recapitulated these toeprints using only purified eIF4A and drug, showing that they reflect eIF4A/RocA complexes clamped directly onto polypurine motifs, bypassing its canonical recruitment via cap and the eIF4F complex<sup>21</sup> (Extended Data Fig. 9f). RNase I footprinting revealed the full extent of the eIF4A protected region centred on the motif (Fig. 3f and Extended Data Fig. 9g).

These eIF4F-independent eIF4A/RocA complexes directly repress translation. We pre-formed such stable complexes on an mRNA during a pre-incubation with recombinant eIF4A and RocA, and then showed that they repressed its subsequent translation in the absence of free RocA (Fig. 4a). Recombinant forms of eIF4A bearing mutations that disrupt either ATP binding or eIF4G binding still retained the ability to clamp onto polypurine RNA in the presence of RocA (Extended Data Figs 8d–f, h, i and 9h–i) and repress translation from the RNA as strongly as wild-type eIF4A/RocA complex (Extended Data Fig. 9j). Furthermore, supplementation of recombinant eIF4A protein into an *in vitro* translation reaction actually strengthened the repressive effect of RocA (Fig. 4b and Extended Data Fig. 9k), confirming the dominant repressive effect of the eIF4A/RocA complex. In contrast, translation repression by Hipp, which decreases the affinity between eIF4A and

calculated by Fisher's exact test. Bin width is 12 nt. **e**, Western blot of SBP translated from uORF and downstream major ORF in RRL with 0.03 μM RocA treatment. Quantification of bands normalized to long form with DMSO treatment is shown. For gel source data, see Supplementary Fig. 1. **f**, Schematic representation of RocA-mediated translation control. RocA clamps eIF4A onto mRNA by selective affinity enhancement for a polypurine motif in eIF4F-, cap-, and ATP-independent manners, which then blocks scanning of pre-initiation complex, introducing premature translation from uORF and inhibiting downstream ORF translation. In **b** and **c**, data represent mean and s.d. (*n* = 3).

RNA and thereby mimics a loss of its function, was relieved by the addition of recombinant eIF4A.

Assembly of an eIF4A/RocA complex could in principle repress 48S formation by blocking 40S attachment to the 5' end of an mRNA or subsequent 43S scanning along the 5' UTR. Because the impact of eIF4A/RocA bound to a single polypurine motif is unaffected by its distance from the 5' end (Extended Data Fig. 9a, l), we infer that eIF4A/RocA bound to these motifs blocks 43S scanning. We also found that eIF4A/RocA could inhibit translation from the polio virus internal ribosome entry site (IRES), which bypasses ordinary 40S recruitment but still depends on scanning (Extended Data Fig. 9a)<sup>25</sup> when we inserted polypurine motifs in the scanned region (Fig. 4c and Extended Data Fig. 9m). Scanning inhibition suffices to explain repression by the eIF4A/RocA complex, although our data do not exclude an additional effect on 40S loading.

Impediments to 43S scanning by stable hairpins<sup>26</sup> or RNA-binding proteins<sup>27</sup> can enhance the translation from upstream open reading frames (uORFs) that otherwise would be skipped. We observed that RocA treatment, but not Hipp treatment, caused an analogous accumulation of translation on 5' UTRs despite the global reduction in footprints on protein-coding sequences (CDSes) (Extended Data Fig. 10a, b). This enhancement occurred specifically on high-sensitivity transcripts (Fig. 4d and Extended Data Fig. 10c). The uORFs activated by RocA showed enrichment of a polypurine motif 20–30 nt downstream of the uORF initiation codon (Fig. 4d, inset), reflecting the distance between the start site and the leading edge of the scanning complex<sup>27</sup>. We tested directly whether eIF4A/RocA complexes on polypurine motifs

can drive cryptic upstream initiation using a reporter mRNA with two alternative start sites that yield distinguishable protein isoforms. Insertion of a polypurine motif 30 nt downstream of the earlier AUG increased translation initiation from this codon upon RocA treatment (Fig. 4e), confirming that clamped eIF4A/RocA complexes on polypurine motifs drive upstream translation initiation. We found evidence that this enhanced upstream initiation could contribute to eIF4A/RocA-mediated repression of downstream CDSes<sup>28</sup>, as RocA-sensitive transcripts showed more pre-existing uORF initiation<sup>29</sup> (Extended Data Fig. 10d, e).

We have shown that RocA induces ATP-independent clamping of eIF4A onto polypurine sequences, creating an inhibitory roadblock for the scanning ribosome (Fig. 4f). Our identification of the eIF4A/RocA binding motif provides the first observation of a drug that stabilizes sequence-selective RNA–protein interactions<sup>30</sup>. RocA may bind near the RNA interface on the N-terminal domain of eIF4A<sup>6</sup>, raising the possibility that the drug directly contacts purine bases of target RNAs. Alternatively, RocA might induce a conformational change leading to direct or indirect recognition of the polypurine motif by protein residues. Future structural insight into this polypurine selectivity may enable rocatate derivatives with altered base selectivities that target different mRNA.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 2 June 2015; accepted 5 April 2016.**

**Published online 15 June 2016.**

- Cencic, R. *et al.* Antitumor activity and mechanism of action of the cyclopenta[*b*]benzofuran, silvestrol. *PLoS ONE* **4**, e5223 (2009).
- Bordeleau, M. E. *et al.* Therapeutic suppression of translation initiation modulates chemosensitivity in a mouse lymphoma model. *J. Clin. Invest.* **118**, 2651–2660 (2008).
- Liu, T. *et al.* Synthetic silvestrol analogues as potent and selective protein synthesis inhibitors. *J. Med. Chem.* **55**, 8859–8878 (2012).
- Santagata, S. *et al.* Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. *Science* **341**, 1238303 (2013).
- Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).
- Sadlish, H. *et al.* Evidence for a functionally relevant rocatamide binding site on the eIF4A–RNA complex. *ACS Chem. Biol.* **8**, 1519–1527 (2013).
- Ingolia, N. T., Ghaemmamghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
- Ingolia, N. T. *et al.* Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Reports* **8**, 1365–1379 (2014).
- Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731–745 (2009).
- Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev.* **12**, 67–83 (1998).
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* **505**, 701–705 (2014).
- Bordeleau, M. E. *et al.* Functional characterization of IRESes by an inhibitor of the RNA helicase eIF4A. *Nature Chem. Biol.* **2**, 213–220 (2006).
- Lindqvist, L. *et al.* Selective pharmacological targeting of a DEAD box RNA helicase. *PLoS ONE* **3**, e1583 (2008).
- Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).
- Thoreen, C. C. *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).
- Lambert, N. *et al.* RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell* **54**, 887–900 (2014).
- Linder, P. & Jankowsky, E. From unwinding to clamping – the DEAD box RNA helicase family. *Nature Rev. Mol. Cell Biol.* **12**, 505–516 (2011).
- Abramson, R. D. *et al.* The ATP-dependent interaction of eukaryotic initiation factors with mRNA. *J. Biol. Chem.* **262**, 3826–3832 (1987).
- König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Struct. Mol. Biol.* **17**, 909–915 (2010).
- Pestova, T. V. & Kolupaeva, V. G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* **16**, 2906–2922 (2002).
- Parsyan, A. *et al.* mRNA helicases: the tacticians of translational control. *Nature Rev. Mol. Cell Biol.* **12**, 235–245 (2011).
- Pause, A. & Sonenberg, N. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.* **11**, 2643–2654 (1992).
- Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J. B. Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–1601 (2013).
- Shirokikh, N. E. *et al.* Quantitative analysis of ribosome-mRNA complexes at different translation stages. *Nucleic Acids Res.* **38**, e15 (2010).
- Balvay, L., Soto Rifo, R., Ricci, E. P., Decimo, D. & Ohlmann, T. Structural and functional diversity of viral IRESes. *Biochim. Biophys. Acta* **1789**, 542–557 (2009).
- Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl Acad. Sci. USA* **87**, 8301–8305 (1990).
- Medenbach, J., Seiler, M. & Hentze, M. W. Translational control via protein-regulated upstream open reading frames. *Cell* **145**, 902–913 (2011).
- Arribere, J. A. & Gilbert, W. V. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res.* **23**, 977–987 (2013).
- Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl Acad. Sci. USA* **109**, E2424–E2432 (2012).
- Pommier, Y. & Marchand, C. Interfacial inhibitors: targeting macromolecular complexes. *Nature Rev. Drug Discov.* **11**, 25–36 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to J. Tanaka for providing hippuristanol, to Y. Tomari for sharing DNA constructs, to H. Asahara and University of California, Berkeley DNA sequencing facility for help with the toeprinting assay, and to A. Pinder and F. Tan for support with deep sequencing analysis. We also thank the members of Ingolia, Lareau, and Tomari laboratories for discussion and technical support. N.T.I. is a Damon-Runyon-Rachleff Innovator supported in part by the Damon Runyon Cancer Research Foundation (DRR-37-15), the Searle Scholars Program (11-SSP-229), and the National Institute of General Medical Sciences of the National Institutes of Health (P50GM102706). This work used the Vincent J. Coates Genomics Sequencing Laboratory at University of California, Berkeley, supported by National Institutes of Health S10 Instrumentation Grants S10RR029668, S10RR027303, and OD018174. S.I. is a recipient of Human Frontier Science Program long-term fellowship. S.N.F. is a Howard Hughes Medical Institute Fellow of the Helen Hay Whitney Foundation.

**Author Contributions** S.I. performed all experiments and analysed the data. Recombinant protein purification and the fluorescence polarization assay were performed with the help of S.N.F. S.I. and N.T.I. designed the experiments and wrote the manuscript.

**Author Information** Sequences generated in this study have been deposited in the Gene Expression Omnibus under accession numbers GSE70211 and GSE79392. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.T.I. ([ingolia@berkeley.edu](mailto:ingolia@berkeley.edu)).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**General methods.** HEK 293 Flp-In T-Rex cells (Invitrogen) were cultured and recombined according to manufacturer's instructions. Stable integrants of SBP-tagged eIF4A were produced by co-transfection of these plasmids along with pOG44 by X-tremeGENE 9 (Roche) and selection using Hygromycin B. RocA, PP242, and thapsigargin were purchased from Sigma. Proteins and DNAs/RNAs were stained with GelCode Blue Stain Reagent (Thermo Scientific) and SYBR Gold Nucleic Acid Gel Stain (Invitrogen), respectively.

**Ribosome profiling.** Library preparation and data analysis were performed according to the method previously described<sup>31</sup>, which monitors mitochondrial ribosomes as well<sup>8,31</sup>. DMSO, RocA, Hipp, and PP242 were added to medium 30 min before cell lysis. The libraries were sequenced on a HiSeq 2000/2500 (Illumina).

**RIP-seq.** Cells with tetracycline-inducible, SBP-tagged eIF4A integrated stably were plated in a 10 cm dish and cultured for 3 days with  $1 \mu\text{g ml}^{-1}$  tetracycline, incubated with DMSO, 0.03  $\mu\text{M}$ , or 0.3  $\mu\text{M}$  RocA for 30 min, washed once with 5 ml of ice-cold PBS, lysed with 600  $\mu\text{l}$  of lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, and 1 mM DTT) containing 1% Triton X-100 and Turbo DNase I (Invitrogen) 25 U ml<sup>-1</sup>, and then clarified by centrifugation for 10 min at 20,000g, 4°C. The supernatant was incubated with 60  $\mu\text{l}$  of Dynabeads M-270 Streptavidin (Invitrogen) equilibrated with lysis buffer containing 1% Triton X-100 at 4°C for 30 min. The beads were washed five times with lysis buffer containing 1% Triton X-100 and 1 M NaCl. SBP-eIF4A and bound RNAs were eluted with 25  $\mu\text{l}$  of lysis buffer containing 5 mM biotin at 4°C for 30 min. All buffers contained 0.001% DMSO with or without 0.03 or 0.3  $\mu\text{M}$  RocA. RNAs were extracted with QIAzol (Qiagen) using the Direct-zol RNA miniprep (Zymo Research). One-third of eluted RNA (~100 ng) was mixed with 1 ng of *in vitro* transcribed, spike-in *Renilla* luciferase RNA (*hRluc*) (see 'DNA constructs') and sequencing libraries were prepared using Tru-seq Ribo-zero gold kit (Illumina). Libraries were sequenced on HiSeq2000/2500 (Illumina) sequencers.

**iCLIP.** Cells were cultured as described in 'RIP-seq'. After medium was substituted with ice-cold PBS, the dishes on ice were irradiated with 150 mJ cm<sup>-2</sup> with UV-C (~254 nm) in UVP CL-1000 (UVP). Lysate was prepared as described in 'RIP-seq'. The lysate from a 10 cm dish (600  $\mu\text{l}$ ) was treated with 0.4 U of RNase I (Epicentre) at 37°C for 3 min. Reaction was quenched by the addition of 10  $\mu\text{l}$  of SUPERase In RNase Inhibitor (Invitrogen), and then incubated with 60  $\mu\text{l}$  of Dynabeads M-270 Streptavidin (Invitrogen) equilibrated with lysis buffer containing 1% Triton X-100 at 4°C for 30 min. The beads were washed by CLIP wash buffer (20 mM Tris-Cl pH 7.4, 1 M NaCl, 2 mM EDTA, 1 mM DTT, and 1% Triton X-100) twice, by CLIP wash buffer containing 0.1% SDS and 0.05% sodium deoxycholate twice, and then by lysis buffer containing 1% Triton X-100 twice. After discarding the supernatant, the beads were incubated with 10 U T4 PNK (NEB), 1  $\times$  PNK buffer, and 0.33  $\mu\text{M}$  <sup>32</sup>P- $\gamma$ [ATP] (3,000 Ci mmol<sup>-1</sup>, PerkinElmer) in 10  $\mu\text{l}$  at 37°C for 5 min and washed once with lysis buffer containing 1% Triton X-100. RNA-crosslinked proteins were eluted by 20  $\mu\text{l}$  of lysis buffer containing 1% Triton X-100 and 5 mM biotin at 37°C for 5 min, run onto NuPAGE (Invitrogen), and transferred to nitrocellulose membrane 0.45  $\mu\text{m}$  (Biorad). The images of <sup>32</sup>P-labelled RNA-protein complex on the membrane were acquired by Typhoon TRIO (Amersham Biosciences). The membrane with the region containing SBP-eIF4A/RNA complexes was excised and treated with 0.1  $\mu\text{g ml}^{-1}$  Proteinase K (Thermo Scientific), 200 mM Tris-Cl, pH 7.4, 25 mM EDTA, pH 8.0, 300 mM NaCl, and 2% SDS in 200  $\mu\text{l}$  at 55°C for 20 min. RNAs were isolated by phenol/chloroform extraction and ethanol precipitation. Library preparation was performed according to the method previously described<sup>31</sup> with the following modifications. As linker DNA, 5'-(Phos)NNNNNNIIITGATCGGAAGAGCACAGCTCTGAA(ddC)-3', where (Phos) indicated 5' phosphorylation and (ddC) indicates a terminal 2', 3'-dideoxycytidine, was used. The capital letter Ns indicate random barcode and the letter Is indicate a sample multiplexing barcode. For multiplexing, linker DNAs containing ATCGT for DMSO replicate number 1, AGTAA for DMSO replicate number 2, CGTAA for RocA 0.03  $\mu\text{M}$ , CTAGA for RocA 0.3  $\mu\text{M}$ , and GATCA for RocA 3  $\mu\text{M}$  in I positions were used, respectively. The linker DNAs were pre-adenylated by 5' DNA adenylation kit (NEB) before the ligation reaction. Instead of gel extraction, unreacted linkers were removed by the treatment of the ligation reaction with 5' deadenylase (NEB) and RecJ exonuclease (Epicentre) at 30°C for 45 min. Reverse transcription was performed with an oligonucleotide 5'-(Phos)NNAGATCGGAAGAGCGTCGTAGGGAAAGAG(iSp18)GTGACTGGAGTTCAGACGTGTGCTC-3', where (Phos) indicates 5' phosphorylation and Ns indicate random barcode. PCR was performed with oligonucleotides, 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCC

TACACGACGCTC-3' and 5'-CAAGCAGAAGACGGCATACGAGATCGT GATGTGACTGGAGTTCAGACGTGTG-3'. Libraries were sequenced on HiSeq4000 (Illumina) sequencers. Random barcode was used to eliminate PCR duplicates in the library.

**Bind-n-Seq.** SBP-tagged eIF4A was purified as described in 'RIP-seq', without DMSO or RocA treatment. The beads tethering SBP-eIF4A were treated with 1  $\times$  Micrococcal Nuclease Buffer (NEB), 0.5  $\times$  lysis buffer, 0.5% Triton X-100, and 200 U  $\mu\text{l}^{-1}$  Micrococcal Nuclease (NEB) in 30  $\mu\text{l}$  at 25°C for 30 min, washed five times with lysis buffer containing 1% Triton X-100, 1 M NaCl, and 5 mM EGTA pH 7.4, and rinsed twice with lysis buffer containing 0.1% Triton X-100. The beads were incubated in lysis buffer containing 0.1% Triton X-100, 2 mM AMP-PNP, 0.33 U  $\mu\text{l}^{-1}$  SUPERase In RNase Inhibitor (Invitrogen), 1  $\mu\text{M}$  N<sub>30</sub> RNA ((N)<sub>30</sub>CTGTAGGCACCATCAAT, where letters in bold type represent DNA sequence for reverse transcription primer hybridization) in 30  $\mu\text{l}$  at 37°C for 30 min, and washed five times with lysis buffer containing 0.1% Triton X-100, 2 mM AMP-PNP, and 0.1% DMSO. SBP-eIF4A/RNA complex was eluted with 30  $\mu\text{l}$  of lysis buffer containing 0.1% Triton X-100, 2 mM AMP-PNP, and 5 mM biotin. DMSO (0.1%) with or without 30 or 300 nM RocA was present in all buffers during the RNA binding reaction, wash, and elution. RNAs were extracted with QIAzol (Qiagen) using the Direct-zol RNA miniprep (Zymo Research) and converted into DNA library as the same method of ribosome profiling<sup>31</sup>. For Bind-n-Seq with ADP + Pi, 2 mM ADP, 2 mM Na<sub>2</sub>HPO<sub>4</sub>, 50  $\mu\text{M}$  N<sub>30</sub> RNA, and 3  $\mu\text{M}$  RocA were used.

The 30-nt randomized RNA followed by 3' DNA sequence for reverse transcription priming was designed to avoid ligation biases and sequencing of contaminating RNA fragments from cells during SBP-eIF4A purification, and to cover the entire sequence with a single 50-bp mode of HiSeq (Illumina) sequencers.

Our read depth (~10<sup>8</sup> reads) is less than the theoretical complexity (4<sup>30</sup> ≈ 10<sup>18</sup>), so that the probability that the same sequence appears multiple times in the library is quite low. Therefore, we assumed that reads with exactly the same sequence and length in the library reflect PCR duplicates and counted them only once. Motif enrichment in the range of interest (4–6 nt) was calculated as the ratio of the motif frequency between libraries<sup>16</sup>.

Spearman's correlation of motif number in 5' UTR versus RIP-seq -fold change caused by RocA treatment was used as motif prediction in RIP-seq. High-scoring motifs were defined as those with enrichment of the prediction or the enrichment is >1.5 s.d. from the mean in RIP-seq and Bind-n-Seq, respectively.

**Data analysis.** The reads were aligned to the hg19 human genome reference and the resulting aligned reads were mapped to University of California, Santa Cruz (UCSC) known reference genes, downloaded from the UCSC genome browser in July 2013. A UCSC bed file of known genes was used for the 5' UTR analysis. For mitochondria footprints alignments, we used the RefSeq genes track corresponding to the mitochondrial chromosome (chrM), downloaded from UCSC genome browser. Specific A-site nucleotides were empirically estimated on the basis of the length of each footprint. The offsets were 14 for 26–29 nt and 15 for 30–31 nt. For mitochondria footprints, they were 9 for 26–27 nt, 11 for 28–29 nt, 12 for 30 nt, 13 for 31 nt, and 18 for 32–34 nt. For mRNA fragments, we used offset 14. For measuring footprint density and mRNA fragments of RIP-seq between samples, we restricted our analysis to genes, which have at least 40 and 100 summed counts in each sample, respectively. For CDSs, the analysis only included the transcript positions beginning 15 codons following the start codon and stopping 5 codons preceding the stop codon. For 5' UTRs, we included the transcript positions from the 5' end of the mRNA until 5 codons preceding the start codon. DESeq<sup>32</sup> was used to calculate relative enrichment of genes in the library, including the mitochondrial footprints and spike-in *hRluc* mRNA counts. The calculated -fold change was re-normalized to the value of the summed mitochondria footprints or the spike-in *hRluc* mRNA fragments.

High-sensitivity messages were defined as transcripts with reduction more than twofold from the median, and with *q* value < 0.01, between 3  $\mu\text{M}$  RocA-treated and untreated cells. Low-sensitivity transcripts are defined as same as high-sensitivity but with accumulation over twofold.

For calculation of  $\Delta G$ , RNALfold (ViennaRNA Package)<sup>33</sup> was run with -L30 -g options on 5' UTR sequences from UCSC foldUtr5 table. The minimum  $\Delta G$  along each 5' UTR was used as a representative free energy value for each gene.

The presence of G-quadruplexes was predicted with RNAfold (ViennaRNA Package).

The Gini differences across 5' UTRs were calculated using published data<sup>11</sup>. Analysis was restricted to the mRNAs bearing 5' UTRs having one or more reads on A/C positions on average.

'uORF translation intensity' was calculated using published data<sup>29</sup>. To incorporate the number and intensity of each upstream initiation site in the 5' UTR, we calculated the density of 5' UTR footprints for each transcript as mentioned

**Code availability.** Scripts to run the analyses mentioned above are available upon request.

For competition assay, the complexes were preformed with 1 mM ATP or AMP-PNP, 1  $\mu$ M eIF4A, 10 nM FAM-labelled RNA, and 50  $\mu$ M RocA and chased with 100  $\mu$ M non-labelled RNA. Because of the low affinity, 50  $\mu$ M eIF4A was used with ATP and DMSO.

***In vitro* translation and toeprinting assay.** *In vitro* translation was performed with nuclease-treated RRL system (Promega), according to the manufacturer's instructions. Reporter mRNAs (50 nM; see 'DNA constructs') was incubated in 50% RRL with RocA (concentration shown in the figure legends) or 1% DMSO in 10 µl at 30 °C for 1 h. For the detection of SBP, 20 µl of the reaction was used with uORF + CAACAA or uORF + AGAGAG mRNAs and concentrated with 10 µl of Dynabeads M-270 Streptavidin (Invitrogen).

Dideoxy-terminated sequencing of RNA was performed by reverse transcriptions using 0.125 mM individual dideoxy-NTP and 0.5 mM each deoxy-NTP with the same 5' 6-FAM-labelled primer and ProtoScript II, according to the manufacturer's instructions.

**Polysome profiling.** Cell lysate was prepared as described previously<sup>31</sup>. Lysate containing 15 µg total RNA was loaded on to 10–50% linear sucrose gradients containing 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 100 µg ml<sup>-1</sup> cycloheximide, and 2 U ml<sup>-1</sup> SUPERase RNase Inhibitor and sedimented by ultracentrifugation at 220,000g for 2 h at 4°C with an SW41 rotor (Beckman Coulter). Gradients were fractionated using Gradient station (Biocomp). Ultraviolet absorbance was detected by ECONO UV monitor (Biorad).

containing 1% Triton X-100, and then clarified by centrifugation with 20,000g at 4 °C for 10 min. Nascent peptides were labelled with 5  $\mu$ M Alexa Fluor 488 Azide (ThermoFisher Scientific) by a Click-it Cell Reaction Buffer Kit (ThermoFisher Scientific) according to the manufacturer's instructions and run on SDS-PAGE. Images were acquired by FluorChem R imaging system (ProteinSimple) and quantified by AlphaView (ProteinSimple).

PTGES3 (uc001slu.4): GCGCGCCGCGCTCACCACCCCTCGTT  
TGCACGCACGCAGCTTATCTCTCGTCTCGCGCCCTTTCTACACT  
TTCTCTTTCTCCCCGACCGGAGGACGCGCTTTCCGCGCGGTGCATTC  
TGGGCCCCGAGGTCGAGGACCCGCGCTGCCCGCTCGCTAGGGAAG  
CGAGAAGAGGCCGCGACCGGAGAGAAAAAGCGAGTCGCCACCGGAG  
AGAAGTCGACTCCCTAGCAGCAGCGCCGCCAGAGAGGCCCGCCAC  
CAGTTCGCGCGTCCCCCTGCCCCGTTACA.

HNRNPC (uc001vzy.3): AGGAATGGGGCGGGGACTAGGCCTT  
CGCCTCGGCGGCAGAGGAGACTCGGGGGCCATTTTGTGAAGAGACGAA  
GACTGAGCGTGTGTGGCCGCGTTGCCGACCTCCAGCAGCAGTCGGCT  
TCTCTACGCACAACCCGGAGGTAGGAGACTCAGAATCGAATCTCTTCT  
CCCTCCCCTCTTGTGAGATTTTGTGACTTTCAGCTACATTTTCGGCT  
TTGTGAGAAACCTTACCATCAAAACAG.

HCV IRES: GCCAGCCCCCTGATGGGGGCGACACTCCACCATGAATC  
ACTCCCCCTGTGAGGAACACTGTCTTACGCGAGAAAGCGTCTAGCCATG  
GCGTTAGTATGAGTGTCTGCGACGCTCCAGGACCCCCCTCCCGGGAG  
AGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGAATTGCGCAGGAG  
GACCGGGTCTTTCTTGAGAGTTACCGCTCAATGCTGTGAGAAATTGGG  
CGTGCCCCCGCAAGACTGCTAGCCGAGTAGTGTTGGGTGCGCAAAGGC  
CTTGTTGGTACTGCTGTATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTC  
GTAGACCGTGCACCATGAGCACAATCCTAAACCTCAAAGAAAAACCA  
AACGTAAC.

G-quadruplex: CTAGGTTGAAAGTACTTTGAC**CGGCGGGCGGTCAA**  
TCTTAC**GGCGGGCGG**ACATAGATA**CGGCGGGCGG**TAGAACTA  
**CGGCGGGCGG**GATTAGAATAGTAAA (where letters in bold type represent  
G-quadruplex-forming sequences).

Randomized control for G-quadruplex: CTAGGGCGCACGTACTT  
CGACAACGTCAGCGTTCAGCGTTCCAACGTCAGCGTACAGCGATCCAA  
CGTCAGCGTTCTGCGCTACAACGTCAGCGTATCCGCGTAGCACA.

CAA repeat: GAACAACAACAACAACAACAACAACAACAACAACAAC  
AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA  
CACC.

7× AGAGAG motifs: GAAAGAGAGCAACAAAGAGAGCAACAAAGAG  
AGCAACAAAGAGAGCAACAAAGAGAGCAACAAAGAGAGCAACAAAGA  
GAGCACC.

1× AGAGAG left: GAAAGAGAGCAACAACAACAACAACAA  
CAACAACAACAACAACAACAACAACAACAACAACAACAACA  
CAACACC.

1× AGAGAG middle: GAACAACAACAACAACAACAACAACAACAACA  
CAACAAAGAGAGCAACAACAACAACAACAACAACAACAACAACA  
CACC.

1× AGAGAG right: GAACAACAACAACAACAACAACAACAACAACA  
CAACAACAACAACAACAACAACAACAACAACAACAACAACAAGAGA  
GCACC.

Polio virus IRES: TTAAACAGCTCTGGGGTGTACCCACCCCAG  
AGGCCCACGTGGCGGCTAGTACTCCGGTATTGCGGTACCCTTGACGC  
CTGTTTTTATACTCCCTTCCCGTAACTTAGACGCACAAAACCAAGTTCAA  
TAGAAGGGGGTACAAACCAAGTACCACCACGAACAAGCATTCTGTTTC  
CCCGGTGATGTCGTATAGACTGCTTGCCTGGTTGAAAGCGACGGATCCG  
TATCCGCTTATGTACTTCGAGAAGCCGATACCACCTCGGAATCTTCG  
ATGCGTTGCGCTCAGCACTCAACCCAGAGTGTAGCTTAGGCTGATGAG  
TCTGGACATCCCTACCGGTGACGGTGGTCCAGGCTGCGTTGGCGGCC  
TACCTATGGCTAACGCCATGGGACGCTAGTTGTGAACAAGGTGTGAAG  
AGCCTATTGAGCTACATAAGAATCTCTCCGGCCCTGAATGCGCGCTAATC  
CCAACCTGGGAGCGGTGGTCACAAACAGTATTGGCTGTGCTGAAC  
GCGCAAGTCCGTGGCGGAACCGACTACTTTGGGTGTCGCTGTTTCCTT

TTATTTTATGTGGCTGCTTATGGTGACAATCACAGATTGTTATCATATAA  
CGCAATTGGATTGGCCATCCGGTGAAAGTGAGACTCATTATCTATCTG  
TTTGCTGGATCCGCTCCATTGAGTGTGTTTACTCTAAGTACAATTTCAAC  
AGTTATTTCAATCAGACAATTGTATCATA.

Polio virus IRES 3× AGAGAG: TAAAAACAGCTCTGGGGTGTGA  
CCCACCCAGAGGCCACGTGGCGGCTAGTACTCCGGTATTGCGGTACC  
CTTGACGCTGTTTATACCTCCCTCCCGTAACCTAGACGCACAAAACC  
AAGTCAATAGAAGGGGTACAAACCACTACCACCAAGCAACAGCACTT  
CTGTTTCCCCGGTGATGTCGTATAGACTGCTTGGCTGGTTGAAAGCGA  
CGGATCCGTTATCCGCTTATGTACTTCGAGAAGCCAGTACCACCTCGG  
AATCTTCGATGCGTTGCGCTCAGCACTCAACCCAGAGTGTAGCTTAG  
GCTGATGAGTCTGGACATCCCTACCGGTGACGGTGGTCCAGGCTGCG  
TTGGCGGCTACCTATGGCTAACGCCATGGGACGCTAGTTGTGAACAA  
GGTGTGAAGAGCCTATTGAGCTACATAAGAATCCTCCGGCCCCCTGAAT  
GCGGCTAATCCCAACCTCGGAGCAGGTGGTCACAAACAGTGATTGG  
CCTGTCGTAACGCGCAAGTCCGTGGCGGAACCGACTACTTTGGGTGTC  
CGTGTTCCTTTTATTTTATTTGTTGGCTGCTTATGGTGACAATCACAGATT  
GTTATCATATAAGCGAATTGGATTGGCCATCCGGTGAAAGTGAGACTCAT  
TATCTATCTGTTTGGCTGGATCCGCTCCATTGAGAGAGTTTACTCTAAGT  
AGAGAGTCAACAGTTATTAGAGAGAGACAATTGTATCATA.

The following DNA fragments, coding *Drosophila msl-2* 5' UTR and SBP, were amplified by PCR and used for *in vitro* transcription template.

uORF + CAACAA: TAATACGACTCACTATAGGGCAGCATAACCATTTG  
TTGATGACTCGAGACCTCTCAAACGTAACCAACAACAAGCAGTGACA  
CCATGGACGAGAAAACCAACCGGCTGGCGGGGAGGCCAGTGGTGGAA  
GGGCTGGCAGGCGAGCTGGAACAGCTGCGGGCCAGACTGGAACACCA  
CCCCAGGGCCAGAGAGAGCCTAGCGGCGGAGGAGACTACAAAGACC  
ATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACA  
CAAGTGATTCTAGCGATCGCTCGAGCCCGGGAATTCGTTTAAACCTA  
GAGCGGCC.

uORF + AGAGAG: TAATACGACTCACTATAGGGCAGCATAACCATTTGT  
GATGACTCGAGACCTCTCAAACGTAACCAACAAGAGAGGCCAGTGACAC  
CATGGACGAGAAAACCAACCGGCTGGCGGGGAGGCCAGTGGTGGAA  
GGCTGGCAGGCGAGCTGGAACAGCTGCGGGCCAGACTGGAACACCACC  
CCCAGGGCCAGAGAGAGCCTAGCGGCGGAGGAGACTACAAAGACCAT  
GACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACA  
AGTGATTCTAGCGATCGCTCGAGCCCGGGAATTCGTTTAAACCTAGA  
GCGGCC.

Reporter RNAs were *in vitro* transcribed, capped, and polyadenylated using a T7-Scribe Standard RNA IVT Kit, a ScriptCap m<sup>7</sup>G Capping System, a ScriptCap 2'-O-Methyltransferase Kit, and A-Plus Poly(A) Polymerase Tailing Kit (CELLSCRIPT). The capping reaction was skipped for polio virus IRES and polio virus IRES 3× AGAGAG reporters.

For the generation of stable cell-lines, PCR products containing CDS regions of *EIF4AI* mRNA and SBP amplified from cDNA from human adult normal brain (Invitrogen) and from pASW<sup>35</sup> (a gift from Y. Tomari), respectively, were inserted into HindIII site in pcDNA5/FRT/TO (Invitrogen) by Gibson assembly (NEB). P159Q, F163L, and Q195E mutations were introduced by site-directed mutagenesis.

For recombinant eIF4A protein expression, PCR products containing CDS regions of *EIF4AI* mRNA were inserted into pHM-GWA<sup>36</sup> to construct pHisMBP-eIF4A. VX<sub>4</sub>GKT (A82V) and D296A-T298K mutations were introduced by site-directed mutagenesis. His-tag, MBP-tag, tobacco etch virus protease cleavage site, and the N-terminal region of eIF4A (1–237) were cloned into pET-28a, to construct pHisMBP-eIF4A (1–237).

**Reporter assay in HEK 293 cells.** Transfections were performed in 24-well dishes with a TransIT-mRNA Transfection Kit (Mirus) according to the manufacturer's instructions, at half scale. Three hours after transfection, RocA was added to the medium, and 9 h after transfection cells were washed with PBS and lysed with Passive lysis buffer (Promega). The luciferase assay was performed with *Renilla*-Glo Luciferase Assay System (Promega) according to the manufacturer's instructions. Luminescence was detected with a GloMax-Multi Jr System (Promega).

For stable cell lines with SBP-tagged eIF4A and its mutants, HEK 293 Flp-In T-Rex cells were cultured for 4 days with 1 µg ml<sup>-1</sup> tetracycline before the experiments. Tetracycline was included in media during experiments.

**Quantitative PCR.** Cell lysate or *in vitro* translation reaction for luciferase assay was treated with 40 U ml<sup>-1</sup> TurboDNase for 10 min on ice, and then RNA was extracted by TRI Reagent (Sigma) and Direct-zol RNA MiniPrep (Zymo Research). Reverse transcriptions were performed with ProtoScript II (NEB) and random

primer mix (NEB) according to the manufacturer's instructions. Quantitative PCR (qPCR) was performed with Fast EvaGreen qPCR Mix (Biotium) in BioRad CFX96 Touch Real Time PCR Detection System (Bio-Rad) with oligonucleotides, 5'-TCGTCCATGCTGAGAGTGTC-3', and 5'-CTAACCTCGCCCTTCTCCTT-3'. RNA from non-transfected cells or *in vitro* translation reaction without the addition of mRNAs was used as qPCR background.

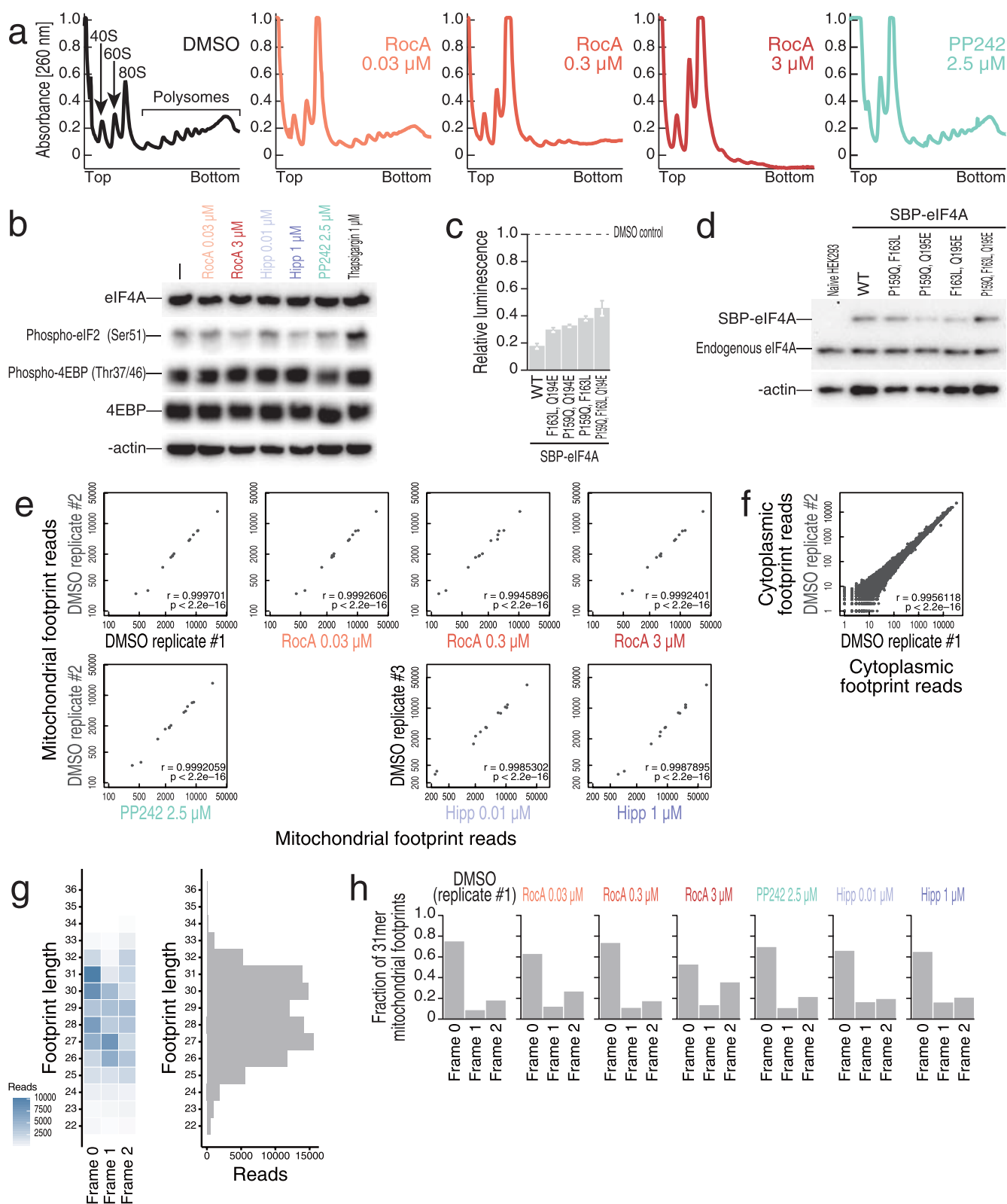
**Purification of recombinant eIF4A proteins.** Typically, BL21 Star (DE3) *Escherichia coli* cells (Invitrogen) transformed with pHisMBP-eIF4A, pHisMBP-eIF4A (VX<sub>4</sub>GKT), pHisMBP-eIF4A (D296A-T298K), or pHisMBP-eIF4A (1–237) in 1.5 L culture were cultivated to an absorbance at 600 nm,  $A_{600\text{ nm}}$ , of 0.5 at 37 °C with 50 µg ml<sup>-1</sup> kanamycin and then grown at 16 °C overnight with 1 mM IPTG. The cell pellets were resuspended in His buffer (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol) with 0.5% NP-40, sonicated, and centrifuged at 35,000g for 20 min. The supernatant was incubated with 1.5 ml bed volume of Ni-NTA Superflow (Qiagen) for 1 h. The beads were loaded on a gravity column and washed with His buffer containing 1 M NaCl. The proteins were eluted with 50 mM Na-phosphate buffer, pH 7.5, 500 mM NaCl, 100 mM Na<sub>2</sub>SO<sub>4</sub>, 250 mM imidazole, 10 mM β-mercaptoethanol, treated with tobacco etch virus protease overnight, dialysed to 20 mM HEPES-NaOH, pH 7.0, 150 mM NaCl, 0.5 mM TCEP, and 10% glycerol, and loaded on MBPTrap HP 5 ml (GE Healthcare). The flow-through fractions were collected, concentrated with Amicon Ultra 10kDa (Millipore), and loaded onto a HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare) equilibrated with 20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 0.5 mM TCEP. The peak fractions were collected, concentrated with Amicon Ultra 10kDa (Millipore), mixed with 0.25 volumes of 80% glycerol, shock-frozen in liquid nitrogen, and stored at –80 °C. All purification steps were performed at 4 °C. Column chromatography was performed using an AKTA purifier (GE Healthcare).

**Pulldown assay.** The lysate of *E. coli* cells expressing eIF4A wild type (WT) or eIF4A D296A-T298K proteins from 1 ml culture was prepared as described in 'Purification of recombinant eIF4A proteins' and incubated with 10 µl of HisPur Ni-NTA Magnetic Beads (Thermo Scientific) at 4 °C for 30 min. The beads were washed five times with His buffer containing 1 M NaCl, rinsed once with 20 mM HEPES-NaOH, pH 7.5, 10 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, and incubated with RRL (Promega) at 25 °C for 30 min. After five washes with His buffer, the proteins were eluted from the beads by SDS sample buffer.

**ATP crosslinking assay.** Recombinant eIF4A WT and VX<sub>4</sub>GKT (10 µM) was incubated with 1 µM [ $\gamma$ -<sup>32</sup>P]-ATP (3,000 Ci mmol<sup>-1</sup>, Perkin Elmer) in 30 mM Hepes-KOH, pH 7.3 (Fisher Scientific), 100 mM KOAc, 5 mM Mg(OAc)<sub>2</sub>, and 1 mM DTT in 20 µl reaction at 37 °C for 15 min. The reactions were exposed to 1500 mJ cm<sup>-2</sup> using UV<sub>254 nm</sub> (CL-1000, UVP) at a distance of 2 cm from the lamp on ice and run on SDS-PAGE. The images were acquired by Typhoon TRIO (Amersham Biosciences).

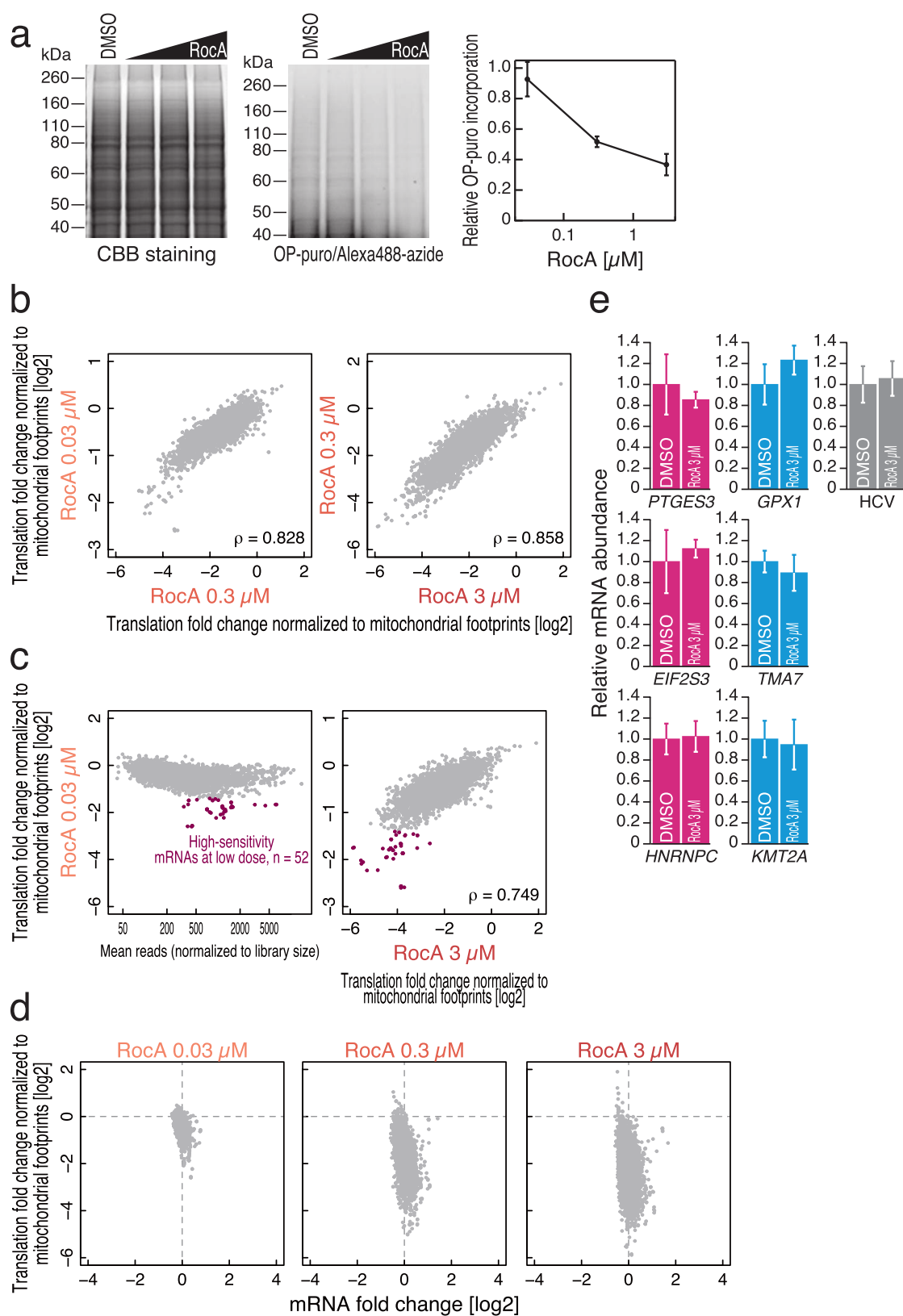
**Western blotting.** Anti-eIF4AI (2490, Cell signaling) (1:1,000), anti-phospho-eIF2α (Ser51) (D9G8 3398, Cell Signaling) (1:1,000), anti-4E-BP1 (9452, Cell Signaling) (1:2,000), anti-phospho-4EBP (Thr37/46) (236B4 2855, Cell Signaling) (1:2,000), anti-β-actin (ab20272, Abcam) (1:1,000), anti-eIF4E (9742, Cell Signaling) (1:1,000), anti-eIF4G (2498, Cell Signaling) (1:1,000), and anti-SBP-tag (SB19-C4 sc-101595, Santa Cruz Biotechnology) (1:1,000) were used as primary antibodies. Chemiluminescence was induced by Pierce ECL Western Blotting Substrate (Thermo Scientific) and images were acquired by a FluorChem R imaging system (ProteinSimple).

- Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols* **7**, 1534–1550 (2012).
- Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
- Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
- Dmitriev, S. E., Pisarev, A. V., Rubtsova, M. P., Dunaevsky, Y. E. & Shatsky, I. N. Conversion of 48S translation preinitiation complexes into 80S initiation complexes as revealed by toeprinting. *FEBS Lett.* **533**, 99–104 (2003).
- Iwasaki, S. *et al.* Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* **39**, 292–299 (2010).
- Busso, D., Delagoutte-Busso, B. & Moras, D. Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal. Biochem.* **343**, 313–321 (2005).
- Galicía-Vázquez, G., Cencic, R., Robert, F., Agenor, A. Q. & Pelletier, J. A cellular response linking eIF4AI activity to eIF4AI transcription. *RNA* **18**, 1373–1384 (2012).



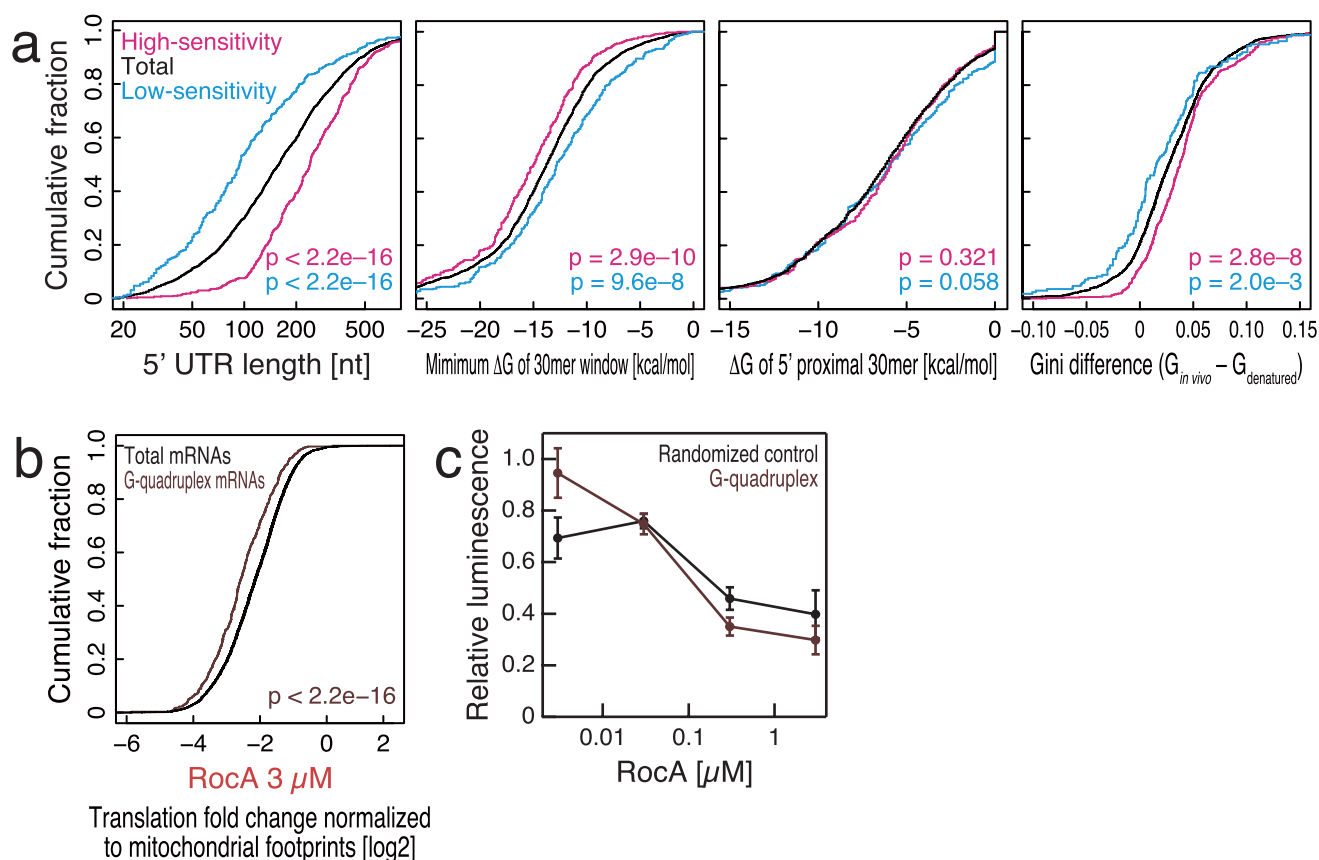
**Extended Data Figure 1 | RocA represses translation, targeting to eIF4A.** **a**, Polysome profiling experiments with RocA and PP242 treatments. RocA disrupts polysomes dose-dependently. **b**, Western blot of phospho-eIF2 $\alpha$  and phospho-4EBP shows that effect of RocA is independent of known translation control targeting to eIFs. Phosphorylation of eIF2 $\alpha$  and dephosphorylation of 4EBP were induced by thapsigargin and PP242, respectively. **c**, **d**, Luciferase reporter assay possessing *PTGES3* 5' UTR (Fig. 1c) with exogenous expression of WT or RocA-resistant eIF4A mutants (c) and western blot of endogenous and exogenous eIF4A (d). eIF4A is the main molecular target of RocA. Data represent mean and s.d. ( $n = 3$ ). **e**, **f**, Correlation of sum of the

footprint reads to 13 mitochondrial mRNAs among different conditions (e) and correlation of sum of the footprint reads from cytoplasmic ribosomes to each transcript between biological replicates (f). Symbol  $r$  is Pearson's correlation.  $P$  value is calculated by Student's  $t$ -test. **g**, **h**, Tile plot of codon periodicity along length of mitochondria footprints (g, left) and mitochondria footprint length distribution (g, right) and codon periodicities of 31-nt mitochondrial footprints among different conditions (h). Footprints with 31-nt length showed most homogenous codon periodicity, and this periodicity was retained with RocA treatment, showing that mitochondrial ribosome translates even in high doses of RocA.



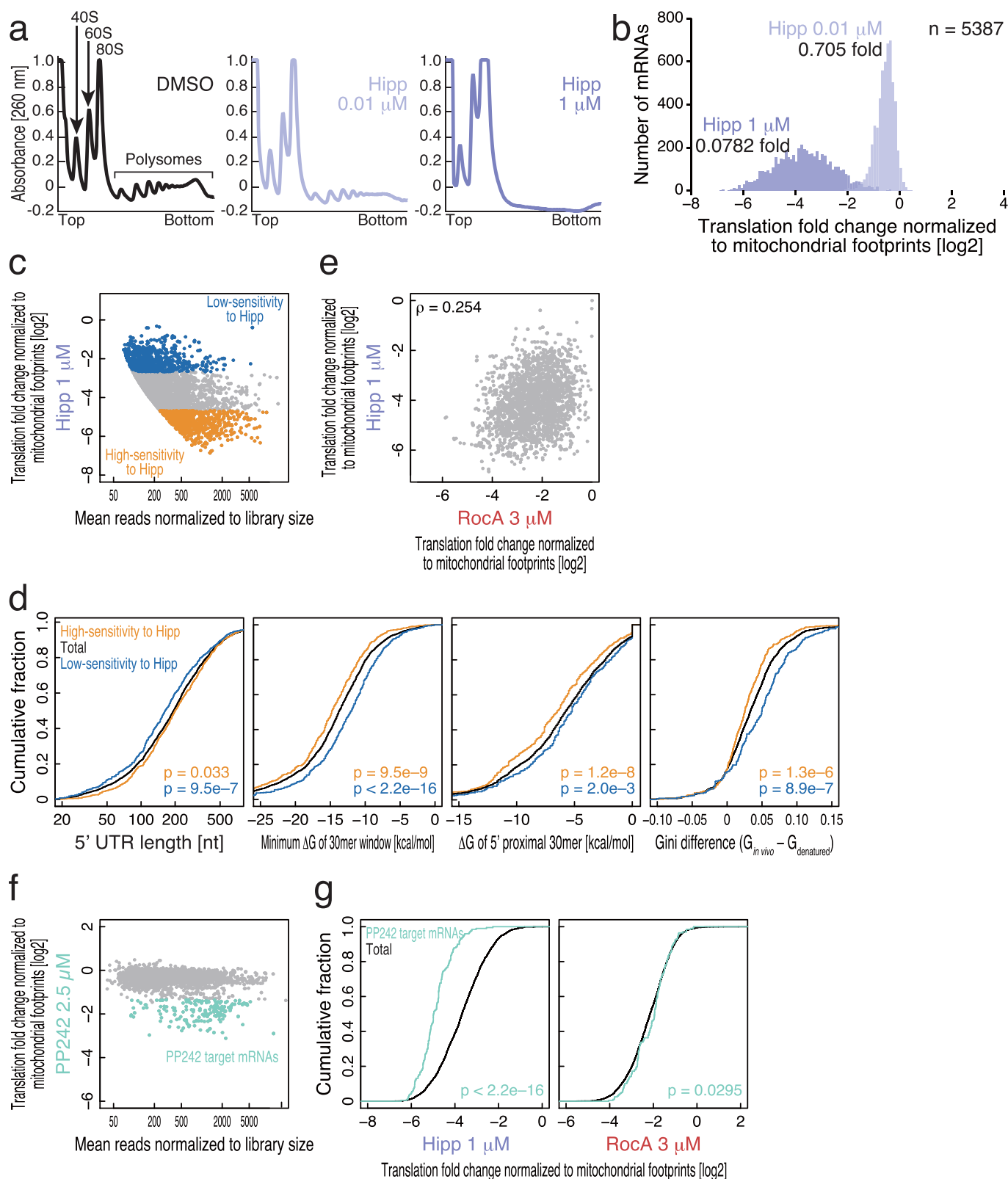
**Extended Data Figure 2 | RocA represses translation without mRNA degradation.** **a**, Metabolic labelling of nascent peptides with OP-puro. The OP-puro incorporated nascent peptides were visualized by Click reaction with Alexa Fluor 488 Azide (middle) and quantified (right). Data represent mean and s.d. ( $n = 3$ ). **b**, Correlation of translation fold change among different concentrations of RocA treatments. **c**, MA plot of mean footprint reads between 0.03  $\mu$ M RocA treatment and non-treatment normalized to

library sizes to footprints fold change by 0.03  $\mu$ M RocA treatment (left) and the correlation of translation fold change between 0.03 and 3  $\mu$ M of RocA treatments (right), highlighting high-sensitivity mRNAs at 0.03  $\mu$ M RocA treatment. **d**, Scatter plots of footprint fold change normalized to mitochondrial footprints and mRNA fold change by RocA treatments. RocA represses translation without significant mRNA change. **e**, qPCR from the samples of Fig. 1c. Data represent mean and s.d. ( $n = 3$ ).



**Extended Data Figure 3 | Secondary structure in 5' UTR is not strong determinant of RocA sensitivity.** **a**, Cumulative fractions along length of 5' UTR, minimum  $\Delta G$  among all 30-mer windows along a 5' UTR,  $\Delta G$  in cap-proximal region (30 nt) of 5' UTR, and Gini difference are plotted to total, RocA high-sensitivity, and RocA low-sensitivity mRNAs. Significance is calculated by Mann–Whitney *U*-test. **b**, Cumulative fractions along translation -fold change by RocA are plotted to total

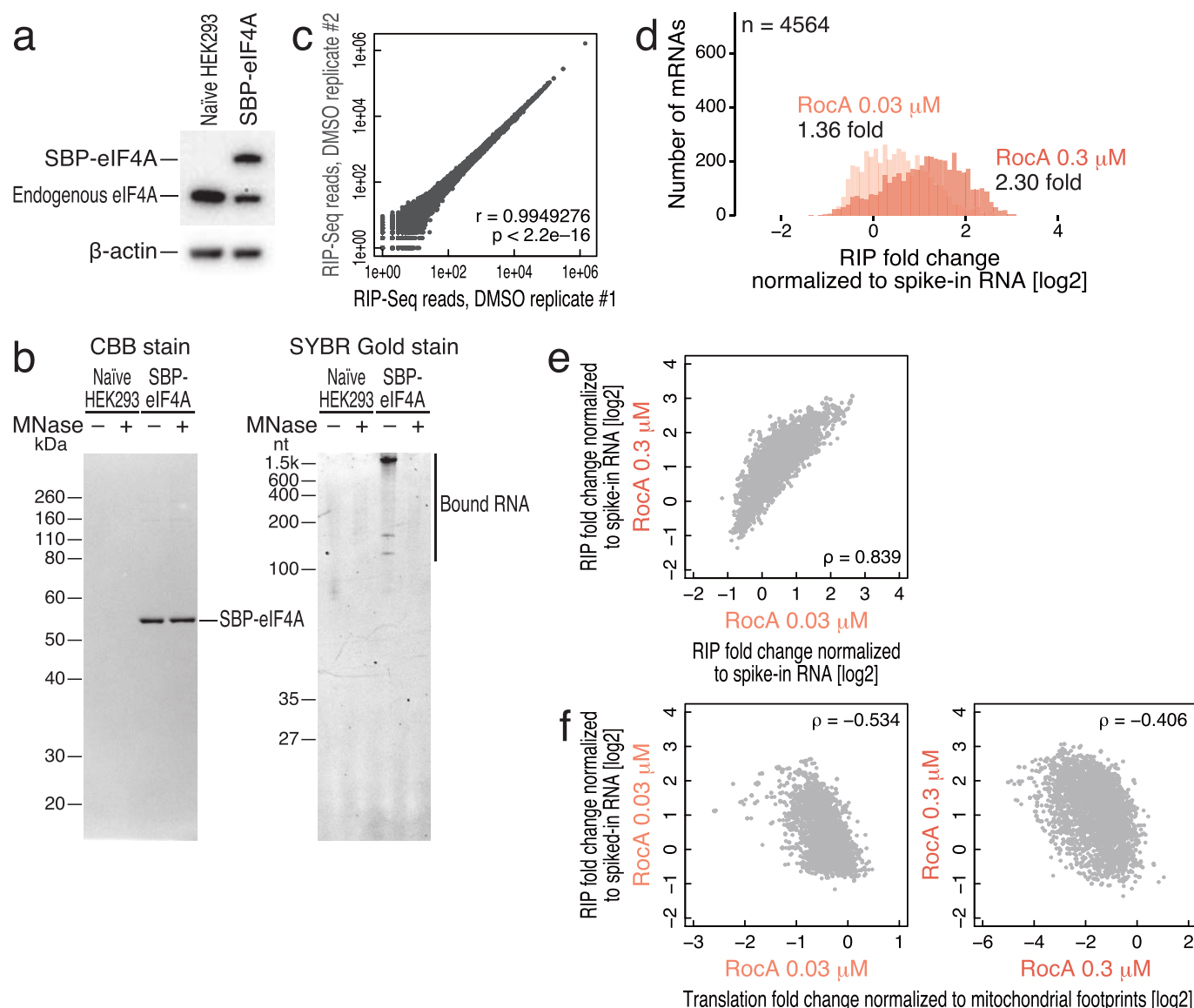
mRNAs and mRNAs with predicted G-quadruplexes in 5' UTRs. Significance is calculated by Mann–Whitney *U*-test. The impact of presence of G-quadruplex in 5' UTR is modest in RocA sensitivity. **c**, The 5' UTRs with G-quadruplexes and randomized control sequence were fused to *Renilla* luciferase and these reporter mRNAs were transfected before treatment with RocA as indicated. Data represent mean and s.d. ( $n = 3$ ). G-quadruplex does not show the prominent RocA sensitivity.



**Extended Data Figure 4 | Characterization of translational inhibition by Hippuristanol and PP242.**

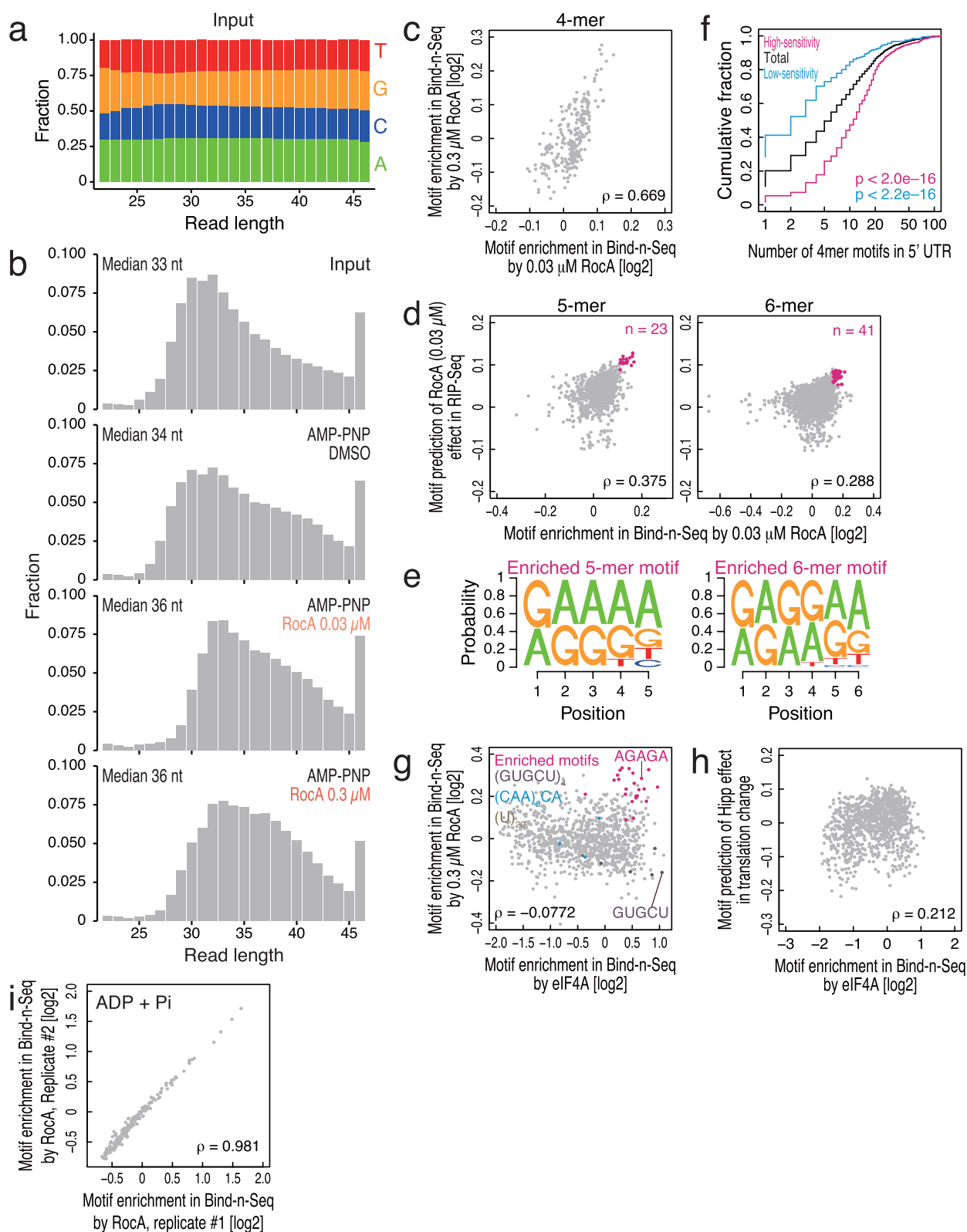
**a**, Polysome profiling experiments with Hipp treatments. Hipp disrupts polysomes dose-dependently. **b**, Histograms of number of transcripts along footprints -fold change with 0.01 and 1  $\mu\text{M}$  Hipp treatment compared with non-treatment, normalized to mitochondrial footprints. Median -fold change is shown. Bin width is 0.1. **c**, MA plot of mean footprint reads between 1  $\mu\text{M}$  Hipp treatment and non-treatment normalized to library sizes to translation -fold change by 1  $\mu\text{M}$  Hipp treatment, highlighting high-sensitivity and low-sensitivity mRNAs. **d**, Cumulative fractions along length of 5' UTR, minimum  $\Delta G$

among all 30-mer windows along a 5' UTR,  $\Delta G$  in cap-proximal region (30 nt) of 5' UTR, and Gini difference are plotted to total, Hipp high-sensitivity, and Hipp low-sensitivity mRNAs. Significance is calculated by Mann-Whitney U-test. **e**, Translation -fold changes by RocA and Hipp are modestly correlated. **f**, MA plot of mean footprint reads between 2.5  $\mu\text{M}$  PP242 treatment and non-treatment normalized to library sizes to translation -fold change by PP242 treatment, highlighting PP242 target mRNAs. **g**, Cumulative distributions of translation -fold change caused by RocA and Hipp treatment are plotted for total and PP242-target mRNAs. Significance is calculated by Mann-Whitney U-test.



**Extended Data Figure 5 | Purification of SBP-tagged eIF4A and co-purified RNA from HEK 293 cells.** **a**, Western blot of exogenous SBP-eIF4A and endogenous eIF4A in tetracycline-inducible stable cell line. Expression of physiological levels of the tagged allele attenuated endogenous eIF4A expression but preserved overall eIF4A levels, probably reflecting the same feedback loop previously reported between eIF4A1 and eIF4A2<sup>37</sup>. **b**, CBB staining of purified SBP-eIF4A and SYBR Gold staining of purified RNA bound to SBP-eIF4A with or without micrococcal nuclease (MNase). **c**, Correlation of sum of the mRNA

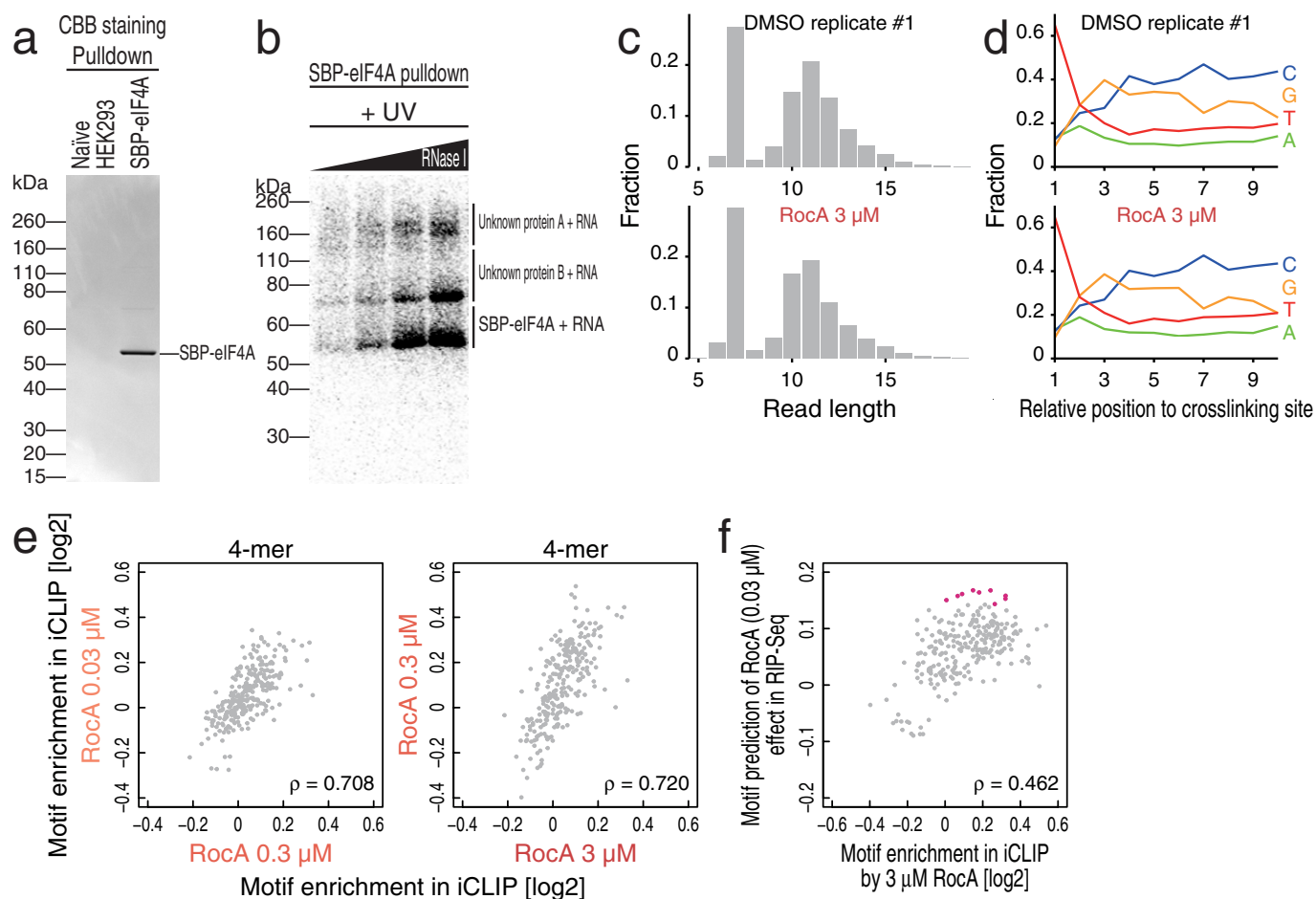
fragment reads of each transcript between biological replicates of RIP-seq.  $P$  value is calculated by Student's  $t$ -test. **d**, Histogram of the number of transcripts along RNA/eIF4A interaction -fold change by RIP-seq when cells are treated with 0.03 or 0.3  $\mu$ M RocA normalized to spiked-in RNA. Data present the same mRNAs analysed in Fig. 1a. Median -fold change is shown. Bin width is 0.1. **e**, Correlation of RIP -fold change between different concentration of RocA treatments. **f**, Correlation of translation -fold change to RIP -fold change with the same concentration of RocA treatment.



### Extended Data Figure 6 | Motif enrichment by Bind-n-Seq.

**a**, Nucleotide composition in each length of reads in input RNAs for Bind-n-Seq. Input RNAs are random in entire read length. **b**, Length distribution of reads from Bind-n-Seq. RNAs bound to eIF4A showed longer length distribution, indicating that eIF4A has preference for longer RNAs. **c**, Correlations of tetramer motif enrichment in Bind-n-Seq by 0.03  $\mu$ M RocA treatment to that by 0.3  $\mu$ M RocA treatment. **d**, Correlations between pentamer and hexamer motif enrichment in Bind-n-Seq by 0.03  $\mu$ M RocA treatment and motif prediction of 0.03  $\mu$ M RocA effect in RIP-seq. **e**, Highest-scoring pentamer and hexamer motifs in Bind-n-Seq and RIP-seq. **f**, Cumulative fractions along number of tetramer motifs (Fig. 2b)

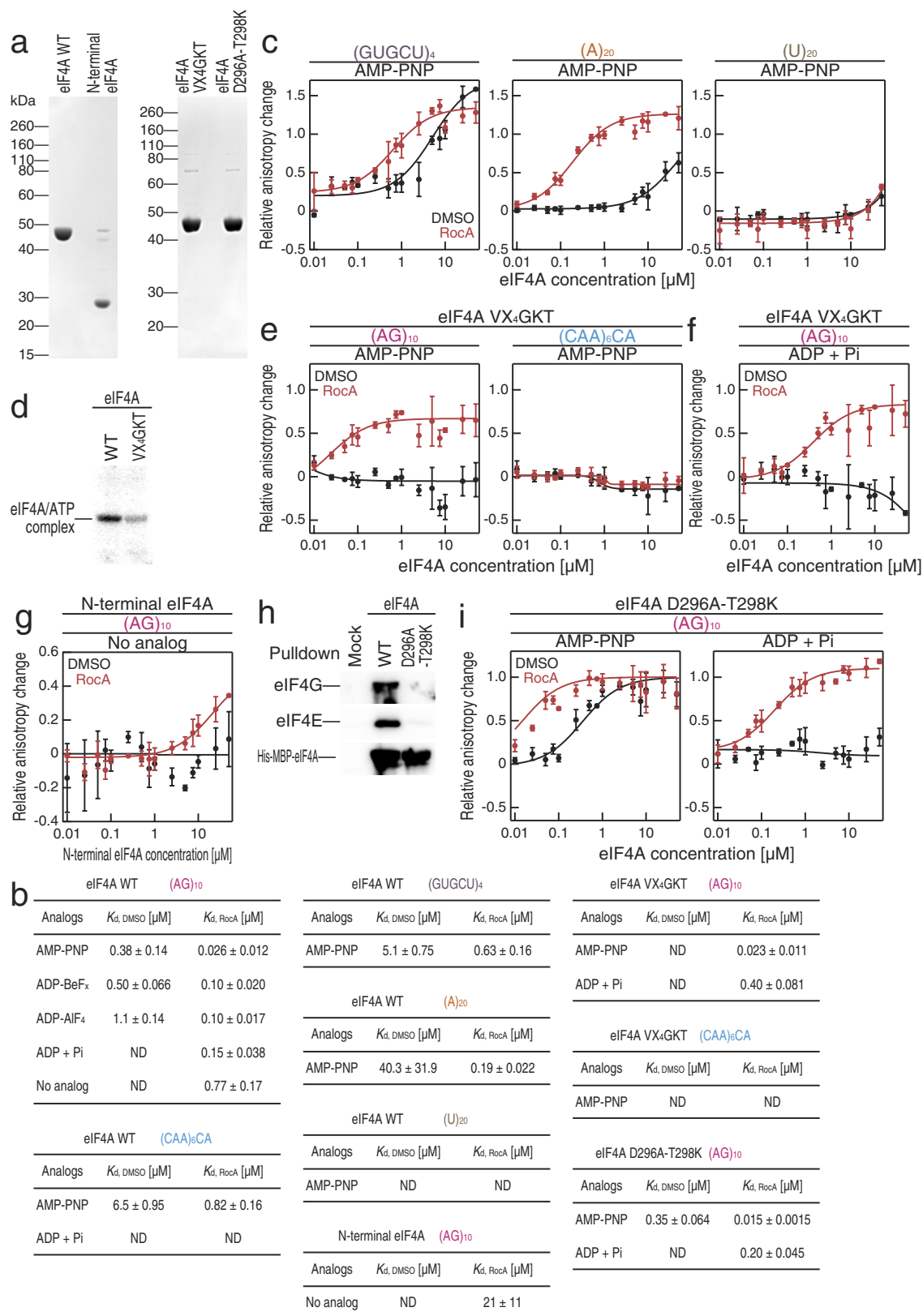
in 5' UTR are plotted to total, RocA high-sensitivity, and RocA low-sensitivity mRNAs. Significance is calculated by Mann-Whitney *U*-test. **g**, Correlations of Bind-n-Seq motif enrichment (pentamer) by eIF4A to that by 0.03  $\mu$ M RocA treatment. The motifs appearing in RNAs used in Extended Data Fig. 8 are highlighted. **h**, Correlation of Bind-n-Seq motif enrichment (pentamer) by eIF4A to motif prediction of Hipp effect in translation change, which is defined as Spearman's correlation of motif number in 5' UTR to translation -fold change by Hipp. mRNAs with high-affinity motif to eIF4A in 5' UTR are resistant to Hipp treatment. **i**, The correlation between enriched motifs of replicates in Bind-n-Seq with ADP + Pi.



#### Extended Data Figure 7 | Characterization of iCLIP data.

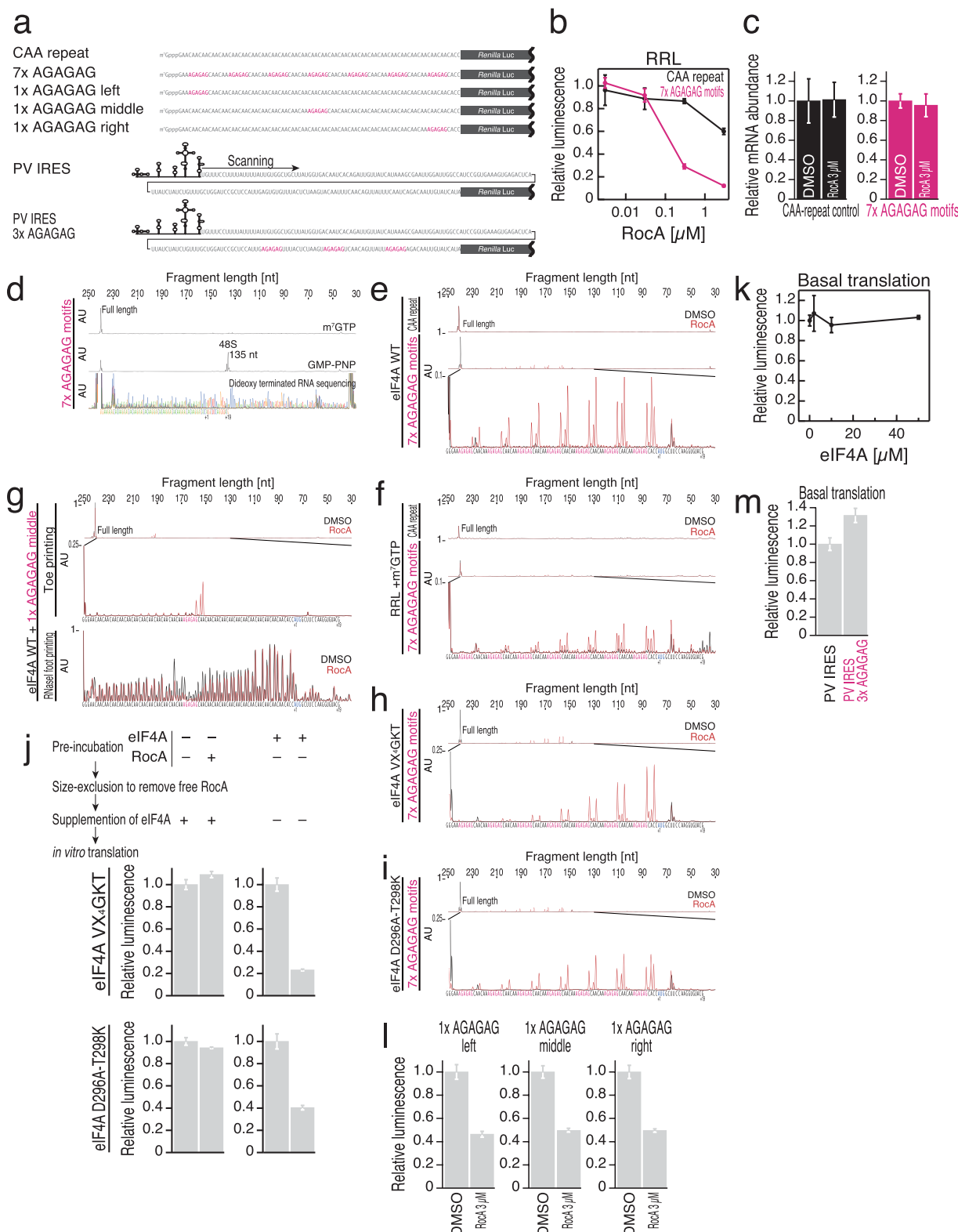
**a**, CBB staining of purified SBP-eIF4A protein in iCLIP procedure. **b**, Visualization of RNA-crosslinked with SBP-eIF4A and unknown proteins by  $^{32}$ P labelling of RNA. We avoided the contamination of RNAs cross-linked to the additional, co-purifying, unknown proteins. **c**, Distribution of read length in iCLIP libraries. Avoidance of contaminating RNAs restricted us to short RNAs, which probably

correspond to the region of RNA physically protected by eIF4A binding, or footprint. **d**, Nucleotide bias along the reads in iCLIP libraries. The crosslinking bias for U may underestimate the preference for polypurine motifs. **e**, Correlations of iCLIP motif enrichment (tetramer) by different RocA concentrations. **f**, Correlations of iCLIP motif enrichment (tetramer) by 3  $\mu$ M RocA and motif prediction of 0.03  $\mu$ M RocA effect in RIP-seq. The motifs shown in Fig. 2a are highlighted.



**Extended Data Figure 8 | eIF4A/RNA affinity measured by fluorescence polarization.** **a**, CBB staining of recombinant proteins used in this study. **b**, Summary of  $K_d$  between RNA and eIF4A among the conditions assayed. **c**, **e–g**, **i**, Direct measurement of the eIF4A/RNA affinity by fluorescence polarization for eIF4A WT, eIF4A (VX<sub>4</sub>GKT), or eIF4A (D296A–T298K)

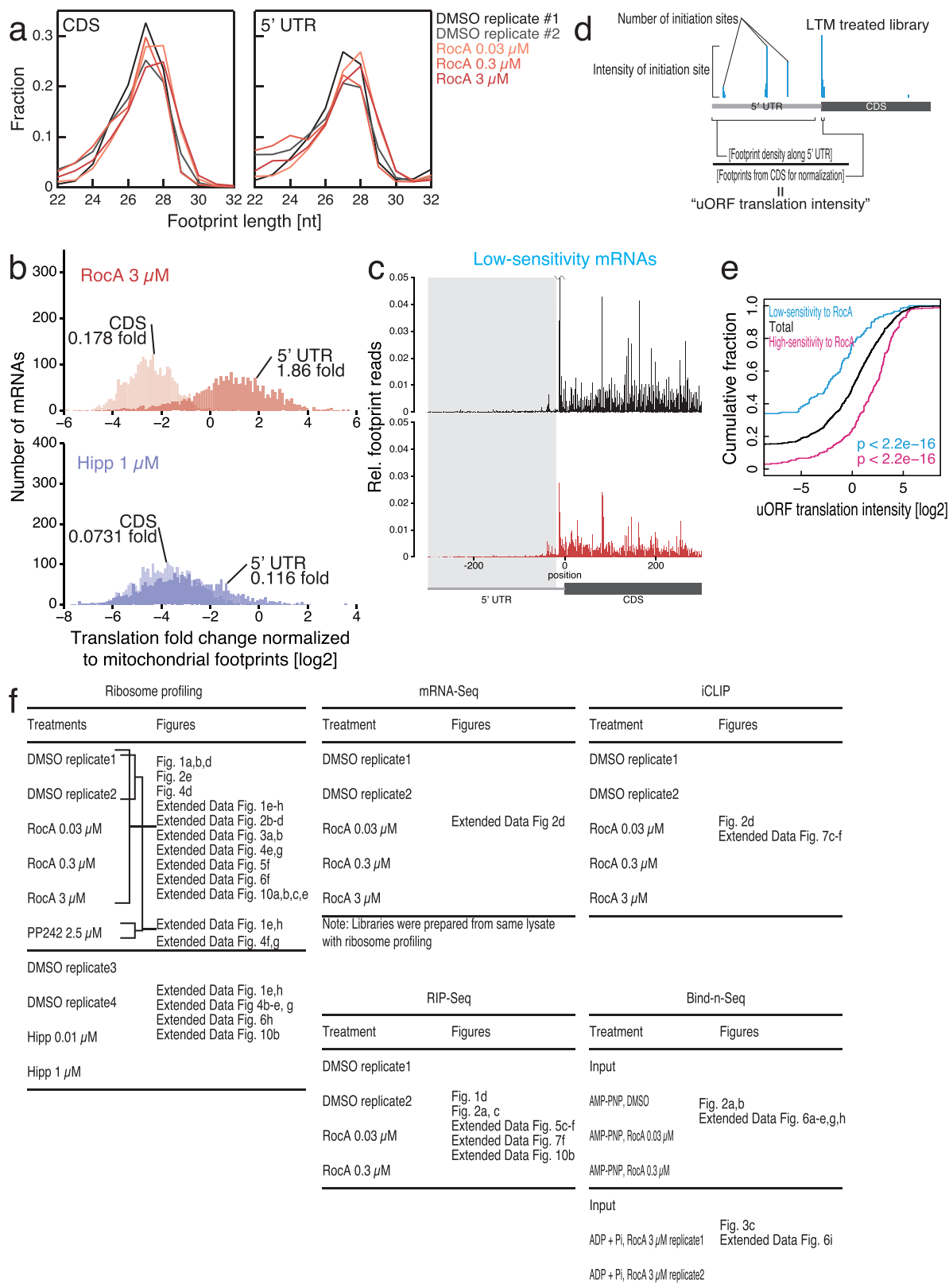
and 5' FAM-labelled RNAs in the presence or absence of RocA. Data represent mean and s.d. ( $n = 3$ ). **d**, ATP crosslinking assay with eIF4A WT and eIF4A (VX<sub>4</sub>GKT). **h**, Pull-down assay with His-MBP-eIF4A expressed in *E. coli* and eIF4E/G in RRL.



### Extended Data Figure 9 | Characterization of toeprinting assay.

**a**, Diagram of the reporters used in this study. **b**, **c**, *In vitro* translation in RRL with mRNAs containing seven polypurine motif (AGAGAG) insertions (**b**) and qPCR from the samples (**c**). **d**, Dideoxy-terminated sequencing of RNA by reverse transcription verified the toeprinting product length terminated by 48S ribosomes. **e**, Ribosome toeprinting assay performed in RRL in the presence of m<sup>7</sup>-GTP in the presence or absence of 3  $\mu$ M RocA treatment. **f**, Toeprinting assay using 10  $\mu$ M recombinant eIF4A in the presence or absence of 10  $\mu$ M RocA treatment. **g**, Toeprinting assay (top) and RNase I footprinting assay (bottom) using 10  $\mu$ M recombinant eIF4A with mRNA containing one AGAGAG motif at the middle in the presence or absence of 10  $\mu$ M RocA treatment.

**h**, **i**, Toeprinting assay using 10  $\mu$ M recombinant eIF4A (VX<sub>4</sub>GKT) or (D296A-T298K) with mRNA containing seven AGAGAG motifs in the presence or absence of 10  $\mu$ M RocA treatment. **j**, Pre-formation of the complex with RocA and eIF4A (VX<sub>4</sub>GKT) or (D296A-T298K) on the mRNA bearing seven polypurine motifs represses the translation from the mRNA in RRL. **k**, Basal translation level from mRNA containing seven AGAGAG motifs with the supplementation of recombinant eIF4A. **l**, *In vitro* translation in RRL with mRNAs with a single polypurine motif (AGAGAG) insertion at the different positions in 5' UTR. **m**, Basal translation level from mRNAs bearing polio virus IRES and polio virus IRES with three AGAGAG motifs. In **b**, **c**, and **h**–**j**, data represent mean and s.d. ( $n = 3$ ).



**Extended Data Figure 10 | The 5' UTR footprints accumulated in RocA treatments come from uORFs.** **a**, The distributions of specific footprint length, which is a hallmark of 80S ribosomes<sup>8</sup>, from CDS and 5' UTR are indistinguishable. **b**, The change in ribosome footprint counts for 5' UTRs and CDSs when cells are treated with 3  $\mu$ M RocA or 1  $\mu$ M Hipp compared with non-treatment, normalized to mitochondrial footprints. Median-fold change is shown. Bin width is 0.1. Analysis is restricted to mRNAs bearing footprints in the 5' UTR in the non-treatment condition. **c**, Meta-gene

analysis of low-sensitivity transcripts to RocA. Reads are normalized to the sum of mitochondrial footprints reads. **d**, The illustration of the definition of uORF translation intensity. **e**, Transcripts sensitive to RocA contain more active uORFs, as measured by cumulative distributions of the uORF translation intensity **c**. Significance is calculated by Mann-Whitney *U*-test. **f**, The summary of deep sequencing-based approaches used in this study and corresponding figures.

# Molecular architecture of the human sperm IZUMO1 and egg JUNO fertilization complex

Halil Aydin<sup>1</sup>, Azmiri Sultana<sup>1</sup>, Sheng Li<sup>2</sup>, Annoj Thavalingam<sup>1</sup> & Jeffrey E. Lee<sup>1</sup>

Fertilization is an essential biological process in sexual reproduction and comprises a series of molecular interactions between the sperm and egg<sup>1,2</sup>. The fusion of the haploid spermatozoon and oocyte is the culminating event in mammalian fertilization, enabling the creation of a new, genetically distinct diploid organism<sup>3,4</sup>. The merger of two gametes is achieved through a two-step mechanism in which the sperm protein IZUMO1 on the equatorial segment of the acrosome-reacted sperm recognizes its receptor, JUNO, on the egg surface<sup>4–6</sup>. This recognition is followed by the fusion of the two plasma membranes. IZUMO1 and JUNO proteins are indispensable for fertilization, as constitutive knockdown of either protein results in mice that are healthy but infertile<sup>5,6</sup>. Despite their central importance in reproductive medicine, the molecular architectures of these proteins and the details of their functional roles in fertilization are not known. Here we present the crystal structures of human IZUMO1 and JUNO in unbound and bound conformations. The human IZUMO1 structure exhibits a distinct boomerang shape and provides structural insights into the IZUMO family of proteins<sup>7</sup>. Human IZUMO1 forms a high-affinity complex with JUNO and undergoes a major conformational change within its N-terminal domain upon binding to the egg-surface receptor. Our results provide insights into the molecular basis of sperm–egg recognition, cross-species fertilization, and the barrier to polyspermy, thereby promising benefits for the rational development of non-hormonal contraceptives and fertility treatments for humans and other mammals.

The journey of a human sperm to an egg ends in the female oviduct, where the active sperm penetrates through the zona pellucida glycoprotein layer of the egg to reach the perivitelline space between the zona layer and the plasma membrane of the oocyte<sup>8–10</sup>. The active sperm then fuses with the oocyte membrane to allow the formation of the zygote<sup>1,4</sup>. At least two membrane-bound proteins, sperm IZUMO1 and egg JUNO, are essential for gamete recognition, fusion, or both<sup>5,6</sup>. Both the *IZUMO1* and *JUNO* (also known as *IZUMO1R*) genes are conserved in other mammals<sup>6,11</sup> (Extended Data Figs 1, 2).

Structural and biochemical studies of IZUMO1 are hampered by difficulties in recombinant protein expression<sup>12</sup>. Using *Drosophila melanogaster* S2 cells, we expressed and purified the extracellular region of human IZUMO1 (residues 22–254) by Ni<sup>2+</sup>-affinity and gel filtration chromatography. Biophysical characterization of IZUMO1 revealed a stable and monomeric protein with extensive mixed  $\alpha$ – $\beta$  secondary structural characteristics (Extended Data Fig. 3). We obtained crystals of unbound IZUMO1<sub>22–254</sub> and determined its structure at 3.1 Å resolution. IZUMO1<sub>22–254</sub> is a monomer and adopts a distinct boomerang shape with dimensions of around 85 Å × 25 Å × 22 Å. The overall structure consists of two domains: a rod-shaped N-terminal four-helix bundle (4HB; residues 22–134) and a C-terminal immunoglobulin-like (Ig-like; residues 167–254) domain (Fig. 1 and Supplementary Fig. 1). Two anti-parallel  $\beta$ -strands ( $\beta$ 1 and  $\beta$ 2) function like a hinge between the 4HB and Ig-like domains.

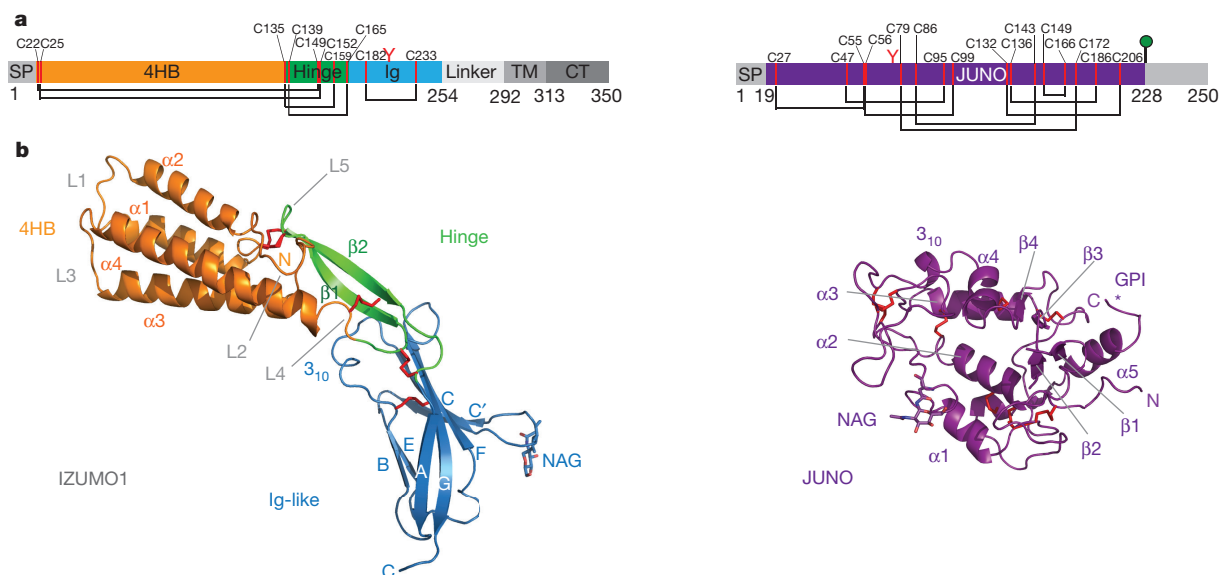
The four helices in the IZUMO1 4HB domain ( $\alpha$ 1,  $\alpha$ 2,  $\alpha$ 3 and  $\alpha$ 4) vary from 14 to 30 residues in length. The helices have amphipathic character with a polar surface exposed to solvent and hydrophobic residues packing into a core. Helices  $\alpha$ 1– $\alpha$ 2 and  $\alpha$ 3– $\alpha$ 4 are connected with short five-residue loops (L1 and L3), and a longer 15-residue loop (L2) links  $\alpha$ 2 to  $\alpha$ 3. The 4HB and hinge regions are stabilized by an extensive network of disulfide linkages (C22–C149, C25–C152, C135–C159, and C139–C165) and charge–charge interactions (H44–D101, E80–K154, and R96–E110) that are conserved in almost all IZUMO1 orthologues and other IZUMO family proteins (Extended Data Fig. 1 and Supplementary Fig. 2).

The IZUMO1<sub>22–254</sub> Ig-like domain resides at the membrane-proximal end of the molecule. It adopts a seven-stranded (A, B, C, C', E, F, G)  $\beta$ -sandwich with the two  $\beta$ -sheets covalently linked with an Ig-superfamily (IgSF) conserved disulfide bond (C182–C233) between strands B and F (Fig. 1). Seven-stranded Ig-like folds classically consist of a 3 + 4 arrangement with  $\beta$ -strands A, B and E forming  $\beta$ -sheet 1, and  $\beta$ -strands C, C', F and G forming  $\beta$ -sheet 2 (ref. 13) (Supplementary Fig. 3). The IZUMO1<sub>22–254</sub> Ig-like domain has a novel 2 + 5 organization representing a previously undescribed IgSF subtype. In IZUMO1<sub>22–254</sub>, strand A interacts with  $\beta$ -sheet 2 rather than  $\beta$ -sheet 1. The disulfide bond preceding  $\beta$ -strand A (C139–C165) may constrain the movement of the strand towards  $\beta$ -sheet 1 and thereby result in this strand switch (Supplementary Fig. 3). We also determined the crystal structure of a slightly longer IZUMO1<sub>22–268</sub> construct at 2.9 Å resolution to gain insights into the C-terminal linker region immediately following the Ig-like domain. The IZUMO1<sub>22–268</sub> structure superimposes well with IZUMO1<sub>22–254</sub> (root mean square deviation (r.m.s.d.) of 1.0 Å over all atoms) (Supplementary Fig. 4). However, no electron density was observed after residue 256, suggesting that the linker region following the Ig-like domain is flexible.

JUNO (previously known as folate receptor- $\delta$  (FOLR- $\delta$ )) is a glycosylphosphatidylinositol (GPI)-anchored, cysteine-rich glycoprotein displayed on the egg surface that has been demonstrated to be the egg receptor of IZUMO1 (ref. 6). We determined the crystal structure of unbound JUNO<sub>20–228</sub> at 1.8 Å resolution. JUNO<sub>20–228</sub> has a globular architecture that is composed of five short  $\alpha$ -helices ( $\alpha$ 1,  $\alpha$ 2,  $\alpha$ 3,  $\alpha$ 4 and  $\alpha$ 5), three  $\beta$ <sub>10</sub> helices and two short two-stranded antiparallel  $\beta$ -sheets (Fig. 1 and Supplementary Fig. 1). Eight conserved disulfide bonds stabilize the core helices  $\alpha$ 1,  $\alpha$ 2,  $\alpha$ 3 and  $\alpha$ 4 and the flexible loops. JUNO shares sequence and structural similarity with human FOLR- $\alpha$  and FOLR- $\beta$  (~58% sequence identity and ~1 Å r.m.s.d. over 197 C $\alpha$  atoms)<sup>14,15</sup> (Extended Data Fig. 4). Despite the close similarities, six key folate binding residues in FOLR- $\alpha$  and FOLR- $\beta$  are not conserved in human JUNO, consistent with previous observations that JUNO does not bind folate<sup>6,14,15</sup> (Extended Data Fig. 4). While this manuscript was under revision, a partial structure of mouse JUNO was determined<sup>16</sup> (Supplementary Fig. 5).

Biolayer interferometry (BLI) and surface plasmon resonance (SPR) were used to measure the binding affinities of human IZUMO1 and

<sup>1</sup>Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada. <sup>2</sup>Department of Medicine, University of California, San Diego, La Jolla, California 92093, USA.

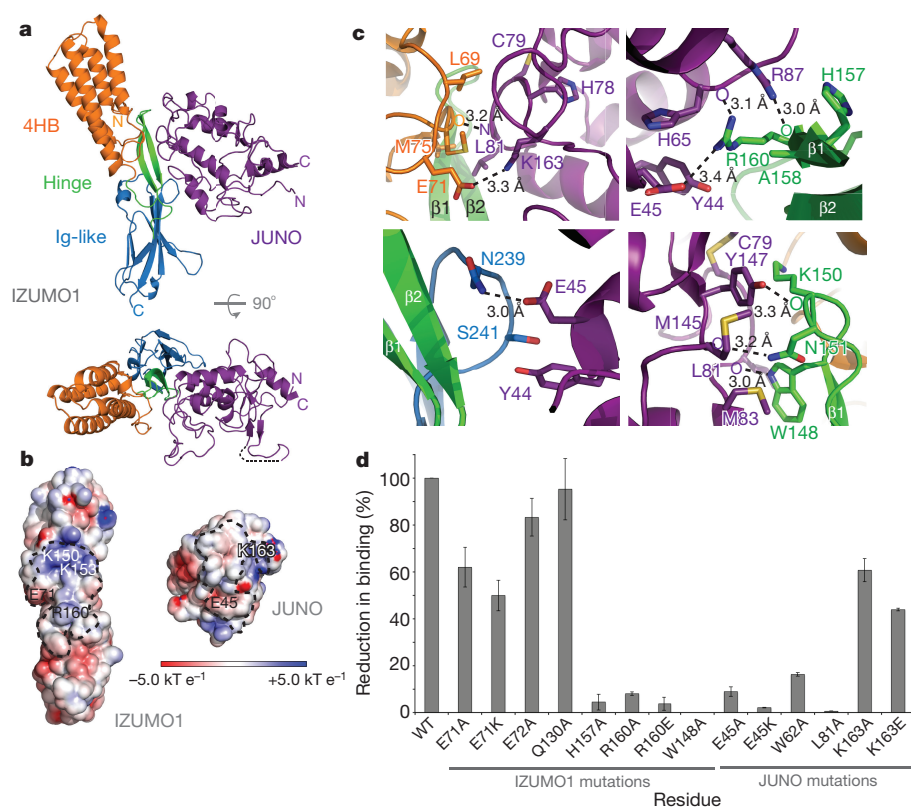


**Figure 1 | Overall structures of human IZUMO1 and JUNO.** **a**, Domain schematics of human IZUMO1 and JUNO. Red Y-shaped and green lollipop symbols denote N-linked glycans and a glycosylphosphatidylinositol (GPI)-anchor, respectively. Regions not observed in the crystal structure are shaded grey. 4HB, four-helix bundle; CT, cytoplasmic tail; Ig, immunoglobulin-like domain; SP, signal peptide; TM, transmembrane region. **b**, Ribbon representation of unbound IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub>. Cysteine residues that form conserved disulfide linkages are highlighted in red.

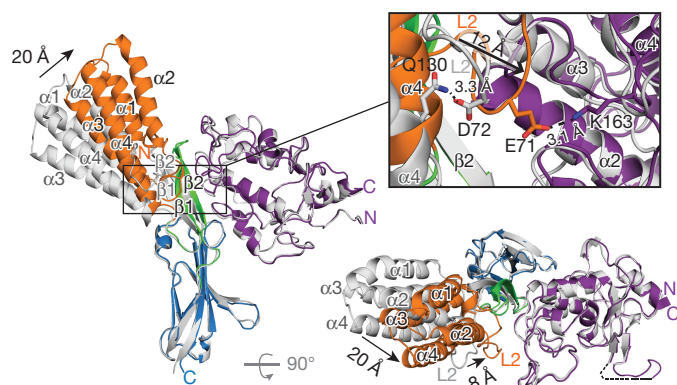
immunoglobulin-like domain; SP, signal peptide; TM, transmembrane region. **b**, Ribbon representation of unbound IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub>. Cysteine residues that form conserved disulfide linkages are highlighted in red.

JUNO. The interaction is stable with a dissociation constant ( $K_d$ ) of about 48–60 nM (Extended Data Fig. 3). In addition, IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> co-purify as a 1:1 complex on size-exclusion chromatography. Our results indicate that these two proteins form a stable complex during the gamete fusion process. To understand the precise molecular interactions of IZUMO1 and JUNO, we determined the structure of the IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex at 2.4 Å resolution. In the asymmetric unit, we observed one IZUMO1<sub>22–254</sub>

molecule binding to one JUNO<sub>20–228</sub> molecule with an interface that spanned a surface area of about 910 Å<sup>2</sup> (Fig. 2 and Extended Data Fig. 5). IZUMO1 has been shown to interact with JUNO via its N-terminal domain<sup>17</sup>. The crystal structure indicates that residues from all three IZUMO1 regions (4HB, hinge and Ig-like) contact JUNO<sub>20–228</sub> through extensive non-bonded van der Waals, hydrophobic and aromatic interactions (more than 60% of total interface interactions; Extended Data Fig. 5). There are also two intermolecular salt



**Figure 2 | IZUMO1–JUNO heterotypic assembly.** **a**, Crystal structure of the human IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex shown as a ribbon diagram. IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> are coloured as in Fig. 1. A disordered loop between the  $\beta$ 1 and  $\beta$ 2 strands of JUNO<sub>20–228</sub> is shown by a black dashed line. **b**, Electrostatic potential surface representation of the IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> binding interface. The footprints of the binding interface are shown by black dashed lines. R160 and E71 on IZUMO1 form a salt bridge with E45 and K163 on JUNO, respectively. **c**, Binding site interactions of IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub>. Side chains of key residues involved in hydrogen bond or salt bridge interactions are shown. **d**, BLI binding affinity analysis of IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> interface mutants. The wild-type IZUMO1–JUNO interaction is normalized at 100% and the binding affinities ( $K_d$ ) for each mutant are shown as percent reductions compared to wild type. All experiments were performed with technical triplicates ( $n = 3$ ), with mean  $K_d$  values  $\pm$  s.e.m. shown in Extended Data Table 1.



**Figure 3 | Conformational changes in IZUMO1 upon binding to JUNO.**

Superposition of structures of unbound IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> (shown in grey) on the IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex (coloured as in Fig. 1). Black arrows highlight the positional changes in secondary structure with the corresponding distances shown in angstroms. Inset, conformational changes within the L2 region during formation of the complex. The L2 region residue D72 and  $\alpha 4$  helix residue Q130 (both shown in grey) form a hydrogen bond in the unbound IZUMO1<sub>22–254</sub> structure. Upon binding to JUNO<sub>20–228</sub>, the L2 region E71 (orange) forms an electrostatic interaction with JUNO<sub>20–228</sub> K163 (purple).

bridges (K163<sub>JUNO</sub>–E71<sub>IZUMO1</sub> and E45<sub>JUNO</sub>–R160<sub>IZUMO1</sub>) and eight hydrogen bond interactions at the interface (Fig. 2 and Extended Data Fig. 5). However, all of these interactions are more than 3.0 Å apart, suggesting that they are weak.

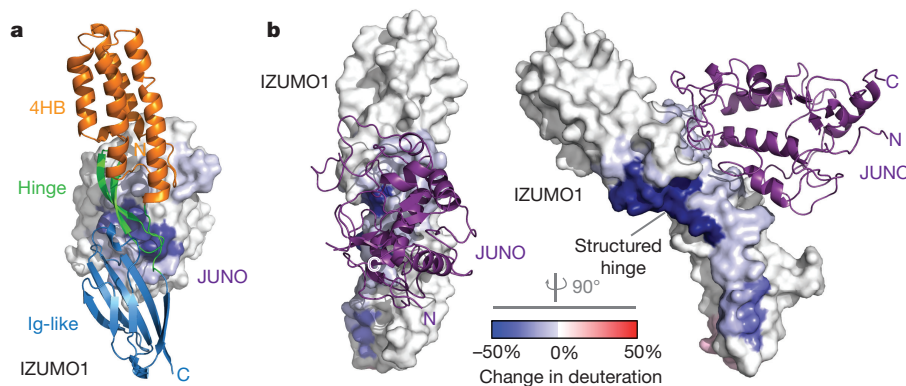
Structural comparison of IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> alone with their bound states revealed no major structural changes in JUNO<sub>20–228</sub> or the Ig-like domain of IZUMO1 upon complex formation (Fig. 3). A major conformational change was observed in the 4HB and hinge regions. All four helices in the 4HB region move about 20 Å towards JUNO<sub>20–228</sub> upon binding, whereas the L2 and hinge regions shift by about 8 Å. As a result, bound IZUMO1<sub>22–254</sub> abandons its distinct boomerang shape and adopts an upright conformation. The structural constraints imposed by the short loops between the  $\alpha 1$ – $\alpha 2$  and  $\alpha 3$ – $\alpha 4$  helices allow the 4HB domain to translate as a single unit (Fig. 3).

To understand the conformational dynamics of the IZUMO1–JUNO interaction in solution, we used a hybrid approach that combined small angle X-ray scattering (SAXS) and deuterium exchange mass spectrometry (DXMS). *Ab initio* SAXS reconstructions of unbound IZUMO1<sub>22–254</sub> revealed a distinct boomerang shape, similar to its crystal structure (Extended Data Fig. 6). When it binds JUNO<sub>20–228</sub>, IZUMO1<sub>22–254</sub> adopts an upright conformation. Our DXMS studies revealed that the residues lining the binding interface exhibited a reduced exchange profile in the complex compared to the unbound state (Fig. 4, Extended Data Fig. 6 and Supplementary

Fig. 6). Moreover, DXMS experiments performed on IZUMO1 alone indicated a high level of exchange in the hinge region, suggesting dynamic flexible motion within this region. Upon binding JUNO, deuterium exchange of residues 127–140 of IZUMO1 in the hinge region was reduced by over 50%, which is greater than the reduction observed in residues at the IZUMO1–JUNO interface (Fig. 4). The strong level of hydrogen/deuterium protection is due to the formation of 10 additional main-chain hydrogen bonds within residues 127–140 of the hinge region. This suggests that the IZUMO1 hinge region is stabilized in a ‘locked’ upright position in the presence of JUNO.

Mutational studies at the IZUMO1–JUNO interface revealed the structural determinants required for binding (Fig. 2, Extended Data Table 1 and Supplementary Fig. 7). Upon binding to JUNO<sub>20–228</sub>, a D72–Q130 hydrogen bond between IZUMO1 L2 and the 4HB  $\alpha 4$  helix is disrupted to form a new intermolecular salt bridge (E71<sub>IZUMO1</sub>–K163<sub>JUNO</sub>) at the interface (Fig. 3). Mutations of IZUMO1 D72 and Q130 to alanine did not affect binding to JUNO, as these residues are not involved at the interface. Alanine and charge-reversal mutations of the intermolecular E71<sub>IZUMO1</sub>–K163<sub>JUNO</sub> salt bridge reduced the  $K_d$  by approximately twofold, suggesting that this ion pair has a minor role in binding. Mutations to the R160<sub>IZUMO1</sub>–E45<sub>JUNO</sub> intermolecular salt bridge result in a roughly 50-fold reduction in binding affinities, suggesting that this second electrostatic interaction has a major role in IZUMO1–JUNO recognition. Using SAXS, we characterized the binding mode of IZUMO1 and JUNO salt bridge mutants that hindered high-affinity complex formation. The SAXS scattering data and reconstructions of mutant complexes show no major changes compared to the wild-type IZUMO1–JUNO complex, and suggest proper complex formation despite up to 50-fold decrease in affinity (Extended Data Fig. 7). Notable reductions in binding (more than 20-fold) were also observed to result from mutations of IZUMO1 and JUNO residues that are conserved in most mammals (W148<sub>IZUMO1</sub>, H157<sub>IZUMO1</sub>, W62<sub>JUNO</sub> and L81<sub>JUNO</sub>). In fact, a mutation of W148<sub>IZUMO1</sub> to alanine completely abolished IZUMO1–JUNO binding. These results agree well with the cell–oocyte binding assay presented in an accompanying crystal structure of the same complex<sup>18</sup>. Together, the results of our mutagenesis study suggest that the interface is probably stabilized through the combined effects of multiple van der Waals, hydrophobic, aromatic and electrostatic interactions. This allows the IZUMO1–JUNO interface to be resilient to mutations.

Although we observed discernible sequence conservation at the complex interface in both proteins, comparative sequence analysis revealed considerable variations among a number of interface residues. Approximately half of the residues (JUNO: Y44, E45, L58, F77, M83, R87, M145, Y147, K163; IZUMO1: L69, V141, K150, N151, K153, E155, A158, Y163, N239, S241) vary across mammalian species (Extended Data Figs 1, 2). Similar to the species-specific recognition employed



**Figure 4 | Comparative DXMS profiles of human IZUMO1 and JUNO binding.** **a**, **b**, Difference in hydrogen–deuterium exchange upon complex formation mapped onto the molecular surfaces of JUNO<sub>20–228</sub> (**a**) and IZUMO1<sub>22–254</sub> (**b**).

by glycoproteins in the zona pellucida that surrounds the egg, these residues may act to restrict productive IZUMO1 and JUNO binding to a specific pair of species<sup>9,19,20</sup>. For example, in primates, these residues are mainly preserved (Extended Data Figs 1, 2), suggesting that the species-specific diversification of IZUMO1 and JUNO may be restricted to non-primate mammals.

Our structural characterizations indicate that human IZUMO1 does not have properties predictive of viral, intracellular or developmental fusogens, such as influenza A virus haemagglutinin (HA) and *Caenorhabditis elegans* epithelial fusion failure-1 (EFF-1) proteins<sup>21–23</sup> (Extended Data Fig. 8). This suggests that IZUMO1 does not function as a direct fusion protein. At least three different fusion mechanisms are possible (Extended Data Fig. 9). First, IZUMO1 may act as a scaffold to recruit a protein complex that contains or regulates other fusion proteins. The requirement of a multiprotein complex for fusion is not unusual, as some viruses, such as herpes simplex virus-1 or Epstein Barr virus, require the formation of a multicomponent fusion complex<sup>24,25</sup>. Alternatively, Inoue *et al.* proposed that monomeric IZUMO1 on the sperm surface interacts with JUNO<sup>26</sup> and, subsequently, a protein disulfide isomerase facilitates the dimerization of IZUMO1 to allow it to interact with another oocyte receptor to facilitate fusion. Finally, the tight heterotypic interaction between human IZUMO1 and JUNO proteins may be sufficient to bring the sperm and egg membranes into close apposition and thereby lead to fusion. Regardless, the conformational changes within IZUMO1 suggest that receptor adhesion triggers the progression of the 4HB domain to the vicinity of the egg membrane and the conformational switch may be part of the structural changes required for fusion. After fusion, the fertilized egg rapidly sheds JUNO molecules into the perivitelline space<sup>6</sup>. Given our measured tight nanomolar affinities between IZUMO1 and JUNO, shed JUNO may essentially act as a rapid 'sperm-sink' to neutralize incoming acrosome-reacted sperm as an additional block to polyspermy. This process may be analogous to the shedding of viral glycoproteins (for example, Ebola virus shed and soluble glycoprotein), which can act as a decoy to absorb antibody responses<sup>27,28</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 16 December 2015; accepted 20 May 2016.**

**Published online 15 June 2016.**

- Wassarman, P. M., Jovine, L. & Litscher, E. S. A profile of fertilization in mammals. *Nat. Cell Biol.* **3**, E59–E64 (2001).
- Evans, J. P. Sperm-egg interaction. *Annu. Rev. Physiol.* **74**, 477–502 (2012).
- Sutovsky, P. Sperm-egg adhesion and fusion in mammals. *Expert Rev. Mol. Med.* **11**, e11 (2009).
- Klinovska, K., Sebkova, N. & Dvorakova-Hortova, K. Sperm-egg fusion: a molecular enigma of mammalian reproduction. *Int. J. Mol. Sci.* **15**, 10652–10668 (2014).
- Inoue, N., Ikawa, M., Isotani, A. & Okabe, M. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* **434**, 234–238 (2005).
- Bianchi, E., Doe, B., Goulding, D. & Wright, G. J. JUNO is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **508**, 483–487 (2014).
- Ellerman, D. A. *et al.* Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. *Mol. Reprod. Dev.* **76**, 1188–1199 (2009).
- Yanagimachi, R. Fertility of mammalian spermatozoa: its development and relativity. *Zygote* **2**, 371–372 (1994).
- Gupta, S. K. *et al.* Mammalian zona pellucida glycoproteins: structure and function during fertilization. *Cell Tissue Res.* **349**, 665–678 (2012).
- Kim, E. *et al.* Sperm penetration through cumulus mass and zona pellucida. *Int. J. Dev. Biol.* **52**, 677–682 (2008).
- Grayson, P. & Civetta, A. Positive selection and the evolution of *izumo* genes in mammals. *Int. J. Evol. Biol.* **2012**, 958164 (2012).
- Marquardt, T. & Helenius, A. Misfolding and aggregation of newly synthesized proteins in the endoplasmic reticulum. *J. Cell Biol.* **117**, 505–513 (1992).
- Halaby, D. M., Poupon, A. & Mornon, J. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng.* **12**, 563–571 (1999).
- Wibowo, A. S. *et al.* Structures of human folate receptors reveal biological trafficking states and diversity in folate and antifolate recognition. *Proc. Natl Acad. Sci. USA* **110**, 15180–15188 (2013).

- Chen, C. *et al.* Structural basis for molecular recognition of folic acid by folate receptors. *Nature* **500**, 486–489 (2013).
- Han, L. *et al.* Divergent evolution of vitamin B9 binding underlies JUNO-mediated adhesion of mammalian gametes. *Curr. Biol.* **26**, R100–R101 (2016).
- Inoue, N. *et al.* Molecular dissection of IZUMO1, a sperm protein essential for sperm-egg fusion. *Development* **140**, 3221–3229 (2013).
- Ohto, U. *et al.* Structure of IZUMO1–JUNO reveals sperm–oocyte recognition during mammalian fertilization. *Nature* <http://dx.doi.org/10.1038/nature18596> (this issue).
- Monné, M., Han, L., Schwend, T., Burendahl, S. & Jovine, L. Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature* **456**, 653–657 (2008).
- Han, L. *et al.* Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell* **143**, 404–415 (2010).
- Chen, E. H. & Olson, E. N. Unveiling the mechanisms of cell-cell fusion. *Science* **308**, 369–373 (2005).
- Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
- Pérez-Vargas, J. *et al.* Structural basis of eukaryotic cell-cell fusion. *Cell* **157**, 407–419 (2014).
- Subramanian, R. P. & Geraghty, R. J. Herpes simplex virus type 1 mediates fusion through a hemifusion intermediate by sequential activity of glycoproteins D, H, L, and B. *Proc. Natl Acad. Sci. USA* **104**, 2903–2908 (2007).
- Sathiyamoorthy, K. *et al.* Assembly and architecture of the EBV B cell entry triggering complex. *PLoS Pathog.* **10**, e1004309 (2014).
- Inoue, N., Hagihara, Y., Wright, D., Suzuki, T. & Wada, I. Oocyte-triggered dimerization of sperm IZUMO1 promotes sperm–egg fusion in mice. *Nat. Commun.* **6**, 8858 (2015).
- Mohan, G. S., Li, W., Ye, L., Compans, R. W. & Yang, C. Antigenic subversion: a novel mechanism of host immune evasion by Ebola virus. *PLoS Pathog.* **8**, e1003065 (2012).
- Dolnik, O. *et al.* Ectodomain shedding of the glycoprotein GP of Ebola virus. *EMBO J.* **23**, 2175–2184 (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by a CIHR Operating Grant (MOP-115066), an NSERC Discovery Grant (RGPIN 435607-13), an Ontario Early Researcher Award (ER-13-09-116), and a Canada Research Chair to J.E.L. S.L. is supported by grants from NIH (1U19AI117905, R01 GM020501 and R01 AI101436). Support for stipends was provided by University of Toronto and Ontario Graduate Scholarships to H.A. and an NSERC USRA to A.T. We thank W. Houry, T. Moraes and C. Spring for access to circular dichroism, SEC–MALS and SPR systems, respectively. This work is based upon X-ray data collected at beamline 08ID-1 at the Canadian Light Source (CLS) and Structural Genomics Consortium (SGC), and SAXS data acquired at the Advanced Light Source (ALS) SIBYLS beamline 12.3.1. The CLS is supported by NSERC, National Research Council of Canada, CIHR, the Province of Saskatchewan, Western Economic Diversification Canada, and the University of Saskatchewan. The ALS is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy (Office of Basic Energy Sciences) through the Integrated Diffraction Analysis Technologies program (DE-AC02-05CH11231), supported by the DOE Office of Biological and Environmental Research; additional support comes from NIH project MINOS (R01GM105404). We thank F. Azimi, J. Cook, A. Dong and N. Ly for technical support, and E. Ollmann Saphire, A. S. Rocca and G. Bikopoulos for critical reading of the manuscript.

**Author Contributions** H.A. designed the project, performed all the cloning, mutagenesis, expression, purification, biophysical characterization and crystallization experiments, collected synchrotron X-ray diffraction data, and determined the crystal structures and SAXS reconstructions. A.S. assisted with BLI and SPR experiments and provided crystallographic guidance for refinement and validation of the crystal structures. A.T. assisted with protein expression and purification. H.A. prepared the samples and S.L. performed and analysed the DXMS data. J.E.L. supervised the research and assisted with BLI experiments. H.A. and J.E.L. analysed and discussed all results, and wrote the manuscript.

**Author Information** Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes: 5F4E (IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex); 5F4Q (JUNO<sub>20–228</sub>); 5F4T (IZUMO1<sub>22–254</sub>); 5F4V (IZUMO1<sub>22–268</sub>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.E.L. ([jeff.lee@utoronto.ca](mailto:jeff.lee@utoronto.ca)).

**Reviewer Information** Nature thanks K. Melcher, M. Okabe and other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

**Protein expression and purification.** The genes encoding full-length human IZUMO1 (GenBank accession number: NM\_182575, residues 1–350) and human JUNO (GenBank accession number: NM\_001199206, residues 1–250) were codon optimized for expression in *D. melanogaster* and gene synthesized (Integrated DNA Technologies). The DNA sequences encoding the extracellular regions of IZUMO1 (residues 22–254 and 22–268) and JUNO (residues 20–228) with a BiP signal peptide were subcloned into a metallothionein promoter pMT expression vector (Invitrogen) modified with a puromycin selection marker (pMT-puro). All protein constructs contain a thrombin cleavage site and a 10 × His affinity tag at the C terminus. Binding interface mutants of IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> were generated using a modified QuikChange PCR-based site-directed mutagenesis protocol. The resulting wild-type and mutant IZUMO1<sub>22–254</sub>, IZUMO1<sub>22–268</sub> and JUNO<sub>20–228</sub> pMT-puro expression plasmids were stably transfected in *Drosophila* S2 cells (Invitrogen) using Effectene transfection reagent (Qiagen) according to the manufacturer's protocol. (S2 cells have been tested by Invitrogen for contamination of bacteria, yeast, mycoplasma and viruses, and been characterized by isozyme and karyotype analysis.) Briefly, *Drosophila* S2 cells were cultured in Schneider's medium (Lonza) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) plus 1 × antibiotic-antimycotic (Gibco), and propagated at 27 °C. The day before transfection, 3 × 10<sup>6</sup> cells were seeded per well in a 6-well plate (Corning) with 3.0 ml complete growth medium and incubated overnight. On the day of transfection, 2 µg expression plasmid was mixed with the transfection reagents and the transfection complexes were added drop-wise onto the S2 cells. At 72 h post-transfection, the cultured medium was replaced with fresh S2 growth medium supplemented with 6 µg ml<sup>−1</sup> puromycin (Bioshop). Subsequently, S2 cells were gradually adapted to FBS-free Insect-XPRESS growth media (Lonza) with 6 µg ml<sup>−1</sup> puromycin. Stably transfected cells were grown to 1 × 10<sup>7</sup> cells per ml in Insect-XPRESS growth medium using vented 2-l polycarbonate Erlenmeyer flasks (VWR) at 27 °C. Protein expression was induced with 500 µM final concentration of sterile-filtered CuSO<sub>4</sub>. Cultured medium was collected 6-days after induction, clarified by centrifugation at 6750g for 20 min, concentrated and buffer exchanged into Ni-NTA binding buffer (20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole) using a Centrimate tangential flow filtration system (Pall Corp.). All IZUMO1 and JUNO proteins were purified by Ni-NTA metal affinity chromatography. Eluted samples were buffer exchanged into TBS (10 mM Tris-HCl (pH 8.0), 150 mM NaCl) using a PD-10 desalting column (GE Life Sciences) and thrombin (EMD Millipore) digested at 22 °C for 24 h (1:2,000 (w/w) enzyme:protein ratio). The cleaved protein samples were then buffer exchanged to a low pH buffer (10 mM sodium acetate (pH 5.6), 150 mM NaCl) and enzymatically deglycosylated using 100 U endoglycosidase H (New England Biolabs) per mg of IZUMO1 or JUNO at 22 °C for 16 h. To prepare the IZUMO1–JUNO protein complexes, deglycosylated IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> samples were mixed at a molar ratio of 1:1 and incubated at 22 °C for 2 h before size-exclusion chromatography on a custom pre-purified Superdex-200 XK 16/70 column equilibrated with TBS. Peak fractions were pooled and protein concentrations were quantified by measuring A<sub>280</sub>.

**Circular dichroism spectroscopy.** Circular dichroism spectra of human IZUMO1<sub>22–268</sub> were acquired on a Jasco J-810 spectropolarimeter using a 1-mm quartz cuvette (Helma). Circular dichroism measurements were conducted with 50–100 µM protein samples purified in 10 mM potassium phosphate (pH 7.5) and 150 mM NaCl buffer. Wavelength scans were recorded at 25 °C between 190 nm and 250 nm and averaged over five accumulations. Data were converted to mean residue ellipticity and secondary structure content was estimated using the K2D algorithm in the DichroWeb analysis server<sup>29</sup>. Thermal denaturation assays were performed at a wavelength of 222 nm by increasing the temperature from 20 °C to 99 °C and monitoring the change in ellipticity as a function of temperature. The data were baseline corrected with buffer blank, normalized between 0 (folded) and 1 (unfolded) and fit to a nonlinear biphasic sigmoidal curve using GraphPad Prism (GraphPad Software).

**Dynamic light scattering (DLS).** IZUMO1<sub>22–254</sub>, JUNO<sub>20–228</sub>, and IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex samples were prepared in TBS with 2% (v/v) glycerol and concentrated to 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0 and 5.0 mg ml<sup>−1</sup> before the DLS measurements. DLS experiments were performed at 25 °C on a DynaPro Plate Reader II (Wyatt Technology). For each condition, 22 µl of sample was loaded in triplicate onto a black 384-well clear-bottom plate (Greiner). Data acquisition was recorded over 5 s with a total of ten acquisitions for each concentration. The polydispersity and hydrodynamic radius (*R*<sub>H</sub>) of the molecules in solution was calculated using Dynamics (v.7) software (Wyatt Technology).

**Size exclusion chromatography–multiangle light scattering (SEC–MALS).** The oligomeric state of tag-removed, glycosylated IZUMO1<sub>22–268</sub> was assessed

by multiangle light scattering. Monomeric bovine serum albumin (BSA) standard (2 mg ml<sup>−1</sup>; 66,432 Da) dissolved in PBS buffer (10 mM phosphate (pH 7.4), 2.7 mM KCl, 137 mM NaCl) was used to calibrate the MALS detectors. IZUMO1<sub>22–268</sub> was purified on an analytical Superdex-75 10/300 GL size exclusion column equilibrated in PBS buffer to ensure proper monodispersity. Then, 600 µg IZUMO1<sub>22–268</sub> was applied onto a PBS-equilibrated Superdex-200 Increase 10/300 GL size exclusion column in-line with a Viscotek MALS detector (Malvern). The data were processed and weight-averaged molecular mass was calculated using the OMNISEC (v. 5.1) software package (Malvern).

**Biolayer interferometry (BLI).** The binding affinities of IZUMO1<sub>22–254</sub> to JUNO<sub>20–228</sub> were measured by biolayer interferometry using a single-channel BLITz instrument (Pall FortéBio), based on protocols previously described<sup>30</sup>. Briefly, purified wild-type or mutant JUNO<sub>20–228</sub> in PBS buffer was biotinylated using the EZ link sulfo-NHS-LC-biotinylation kit (Thermo Pierce), according to the manufacturer's instructions. Excess biotin reagent was removed by overnight dialysis in PBS. All streptavidin-coated (SA) biosensors were hydrated in BLI rehydration buffer (PBS, 0.5 mg ml<sup>−1</sup> BSA and 0.01% (v/v) Tween-20) for 10 min. Biotinylated JUNO<sub>20–228</sub> (bait) was diluted in BLI kinetics buffer (PBS, 0.1 mg ml<sup>−1</sup> BSA and 0.01% (v/v) Tween-20) to a final concentration of 20 µg ml<sup>−1</sup> and immobilized onto a SA biosensor for 90 s. Multiple concentrations of wild-type or mutant IZUMO1<sub>22–254</sub> (analyte) were prepared in BLI kinetics buffer and association to IZUMO1<sub>22–254</sub> was measured over 90 s at 20 °C. Subsequently, the SA biosensor was immersed into BLI kinetics buffer for 90 s to dissociate the analyte. All experiments were performed in triplicate. Two negative controls were performed: BSA and BLI kinetics buffer only against SA biosensors loaded with biotinylated bait to detect non-specific binding. The data were analysed and sensorgrams were step corrected, reference corrected and fit globally to a 1:1 binding model. The equilibrium dissociation constant (*K*<sub>d</sub>), association (*k*<sub>a</sub>) and dissociation (*k*<sub>d</sub>) rate constants and their associated standard errors were calculated using BLITz Pro data analysis (v. 1.1.0.16) software.

**Surface plasmon resonance (SPR).** The affinities and kinetics of wild-type IZUMO1<sub>22–254</sub> binding to wild-type JUNO<sub>20–228</sub> were assessed by SPR on a Biosensing Instruments BI-4000 system at 20 °C using a CM-dextran sensor chip. Prior to immobilization, pH scouting between pH 4.5 and 6.5 was performed to identify the optimal pH for immobilization. Wild-type JUNO<sub>20–228</sub> was immobilized using a coupling buffer containing 10 mM sodium acetate pH 5.0 onto one of two flow channels using the manufacturer's standard amine-coupling protocol. Association of the wild-type IZUMO1<sub>22–254</sub> analyte (0.75 µM, 0.5 µM, 0.375 µM, 0.25 µM, 0.188 µM, 0.125 µM, 0.0937 µM and 0 µM) was measured at a flow rate of 50 µl min<sup>−1</sup> for 90 s. The second flow cell, containing no bait, was injected with PBS buffer in a serial flow and used as a reference. Subsequently, PBS buffer was injected at a flow rate of 50 µl min<sup>−1</sup> over 180 s to dissociate wild-type IZUMO1<sub>22–254</sub>. The cells were regenerated between two analyte runs using the rapid injection protocol involving 8 cycles of 20-µl injections of 0.01 M NaOH-acetate pH 9.0 followed by an equal volume of 1 × PBS. Measurements were performed in triplicate. The resulting SPR sensorgrams were corrected with the reference and blank (0-µM analyte) curves, and fitted globally with a 1:1 Langmuir binding model using BI-Data Analysis and BI-Kinetic Analysis SPR software.

**Crystallization and X-ray data collection.** Purified IZUMO1<sub>22–254</sub>, IZUMO1<sub>22–268</sub>, JUNO<sub>20–228</sub>, and the IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex were concentrated to 10 mg ml<sup>−1</sup>. All crystallization trials were performed at 22 °C by sitting drop vapour diffusion (300 nl protein and 300 nl mother liquor) in 96-well low profile Intelliplates (Art Robbins) using an Oryx8 protein crystallization robot (Douglas Instruments).

**IZUMO1<sub>22–254</sub> and IZUMO1<sub>22–268</sub>.** Initial sparse matrix screening of IZUMO1<sub>22–254</sub> and IZUMO1<sub>22–268</sub> constructs identified needle-shaped crystals in multiple conditions. IZUMO1<sub>22–254</sub> and IZUMO1<sub>22–268</sub> crystals were manually optimized in 48-well MRC Maxi crystallization plates using 2-µl sitting drops. Larger needle-shaped IZUMO1<sub>22–254</sub> crystals appeared the next day and reached a maximum length of ~250 µm within 3–4 days in 0.07 M sodium acetate (pH 4.6), 5.6% (w/v) PEG 4000 and 30% (v/v) glycerol. Larger IZUMO1<sub>22–268</sub> crystals were more difficult to obtain and required further optimization using random microseed matrix screening (rMMS) with Oryx8 (ref. 31). rMMS led to thicker needle crystals in 0.085 M HEPES sodium salt (pH 7.5), 8.5% (v/v) isopropanol, 17% (w/v) PEG 4000 and 15% (v/v) glycerol. These crystals reached a final length of ~200 µm within 4–5 days. All crystals were cryoprotected and flash-cooled in liquid nitrogen. IZUMO1<sub>22–254</sub> and IZUMO1<sub>22–268</sub> crystals diffracted to Bragg spacings of 3.1 Å and 2.9 Å, respectively, and data sets were remotely collected at the Canadian Light Source (CLS) 08ID-1 beamline (Supplementary Fig. 8).

**JUNO<sub>20–228</sub>.** Rod-shaped JUNO<sub>20–228</sub> crystals were grown in 0.02 M magnesium chloride, 0.1 M HEPES sodium salt (pH 7.5), 22% (w/v) polyacrylic acid 5100. Crystals typically appeared after 3–4 days and reached full size in 1 week. The mother liquor supplemented with increasing amounts of sucrose

(up to 30% (w/v) in final solution) was used as a cryoprotectant before being rapidly cooled in liquid nitrogen. JUNO crystals readily produced Bragg reflections at better than 2.0 Å resolution on a Rigaku FR-E Superbright X-ray generator and Saturn A200 HD CCD detector (Rigaku Corp.), and a 1.8 Å resolution data set was collected at the Structural Genomics Consortium X-ray diffraction facility (Supplementary Fig. 8).

**IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex.** Crystals of the protein complex were grown in sitting drops containing equal volumes (1 µl) of purified protein and crystallant (0.1 M MES (pH 6.5), 20% (w/v) PEG 4000 and 0.6 M NaCl). Crystals were observed after 3 days and matured to full size within a week. Crystals were cryoprotected by sequential soaking in mother liquor with 5%, 10%, 20% and 30% (w/v) sucrose. Crystals were directly immersed into liquid nitrogen and screened at the CLS beamline 08ID-1. A data set was collected from a single IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex crystal diffracting to 2.4 Å resolution (Supplementary Fig. 9).

**Structure determination and refinement.** All diffraction data were integrated and reduced with the XDS program package<sup>32</sup> and scaled using Aimless<sup>33</sup> from the CCP4 program suite. Crystallographic data collection and final refinement statistics are presented in Supplementary Table 1.

**JUNO<sub>20–228</sub>.** The structure of JUNO<sub>20–228</sub> was determined by molecular replacement with Phaser<sup>34</sup> using human folate receptor-α (PDB ID: 4LRH) as the search model. Initial characterization of the JUNO<sub>20–228</sub> X-ray data using phenix.xtriage<sup>35</sup> and DETWIN<sup>36</sup> indicated translational pseudosymmetry<sup>37</sup> (TPS) and near-perfect merohedral twinning<sup>38</sup> with an estimated twin fraction of 0.45 (Supplementary Fig. 8). The twinning fraction was calculated from the cumulative distribution of H<sup>39</sup> and Britton plots<sup>40</sup>, with the twin fractions related by the twin law  $k, h, -l$ . It was necessary to apply the twin law throughout the refinement to further refine the JUNO<sub>20–228</sub> structure using phenix.refine<sup>41</sup>.

**IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex.** The initial phases for the IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex were calculated via molecular replacement with Phaser<sup>34</sup>, using the human folate receptor-α structure (PDB ID: 4LRH) as an initial search model. One clear solution ( $Z = 14.6$ ) was identified. Strong electron density was observed for JUNO<sub>20–228</sub> and the 4HB domain of IZUMO1<sub>22–254</sub>. The poly-alanine chain of the IZUMO1<sub>22–254</sub> was initially traced by a combination of phenix.autobuild<sup>42</sup> and manual building with Coot<sup>43</sup>.

Validation of proper sequence registry was confirmed by locating the sulfur anomalous signals from methionine and cysteine residues. Multi-crystal sulfur anomalous data were collected on native IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex crystals. The X-ray beam was focused to 50 µm and the sulfur anomalous signal was measured at a wavelength of 1.7712 Å using a MarMosaic MX300 CCD detector (Rayonix). All crystals were rod-shaped and >400 µm in length, thus allowing us to translate along the rotation axis to expose a fresh undamaged part of the crystal. 360° of data with a rotation angle of 1.0° per frame were collected for each set before translating to a new part of the crystal. Each data set was processed individually with anomalous signal using XDS<sup>32</sup>. Twenty-four data sets with  $R_{\text{meas}} < 10\%$  were merged together using XSCALE<sup>32</sup> and converted to CCP4 data format using XDSCONV<sup>32</sup>, F2MTZ and CAD<sup>36</sup>. The overall  $R_{\text{merge}}$  and anomalous multiplicity for the merged data set to 2.8 Å resolution were 9.9% and 89.6, respectively. An anomalous difference Fourier electron density map was calculated using PHENIX<sup>35</sup> and confirms the correct location for all 38 protein sulfur sites (Supplementary Fig. 9).

**IZUMO1<sub>22–254</sub> and IZUMO1<sub>22–268</sub>.** The structures of IZUMO1<sub>22–254</sub> and IZUMO1<sub>22–268</sub> were determined by molecular replacement. An initial molecular replacement search using the refined IZUMO1<sub>22–254</sub> structure from the IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex failed, probably because of conformational changes between the 4HB and Ig-like domains. A molecular replacement search was performed first using the IZUMO1 Ig-like domain (residues 167–254) followed by a second molecular replacement search using the 4HB and hinge regions (residues 22–166). Clear solutions were identified for both sections in IZUMO1<sub>22–254</sub> and IZUMO1<sub>22–268</sub>.

All structures were manually rebuilt using Coot<sup>43</sup> and refined using phenix.refine<sup>41</sup>. No non-crystallographic (NCS) symmetry restraints were employed except in the case of JUNO<sub>20–228</sub>, where a four-fold NCS was applied. All β-strands and α-helices were real-space refined with torsional secondary structural restraints using Coot. Torsion-angle simulated annealing refinement, starting at 5,000 K, with individual atomic displacement and Translation/Libration/Screw (TLS) groups was carried out using Phenix. Owing to the lower resolutions of the IZUMO1<sub>22–254</sub> and IZUMO1<sub>22–268</sub> data, these structures were refined with grouped B-factor refinement. Calculation of annealed  $2|F_o| - |F_c|$  composite omit maps<sup>44</sup> helped minimize model bias during rebuilding.

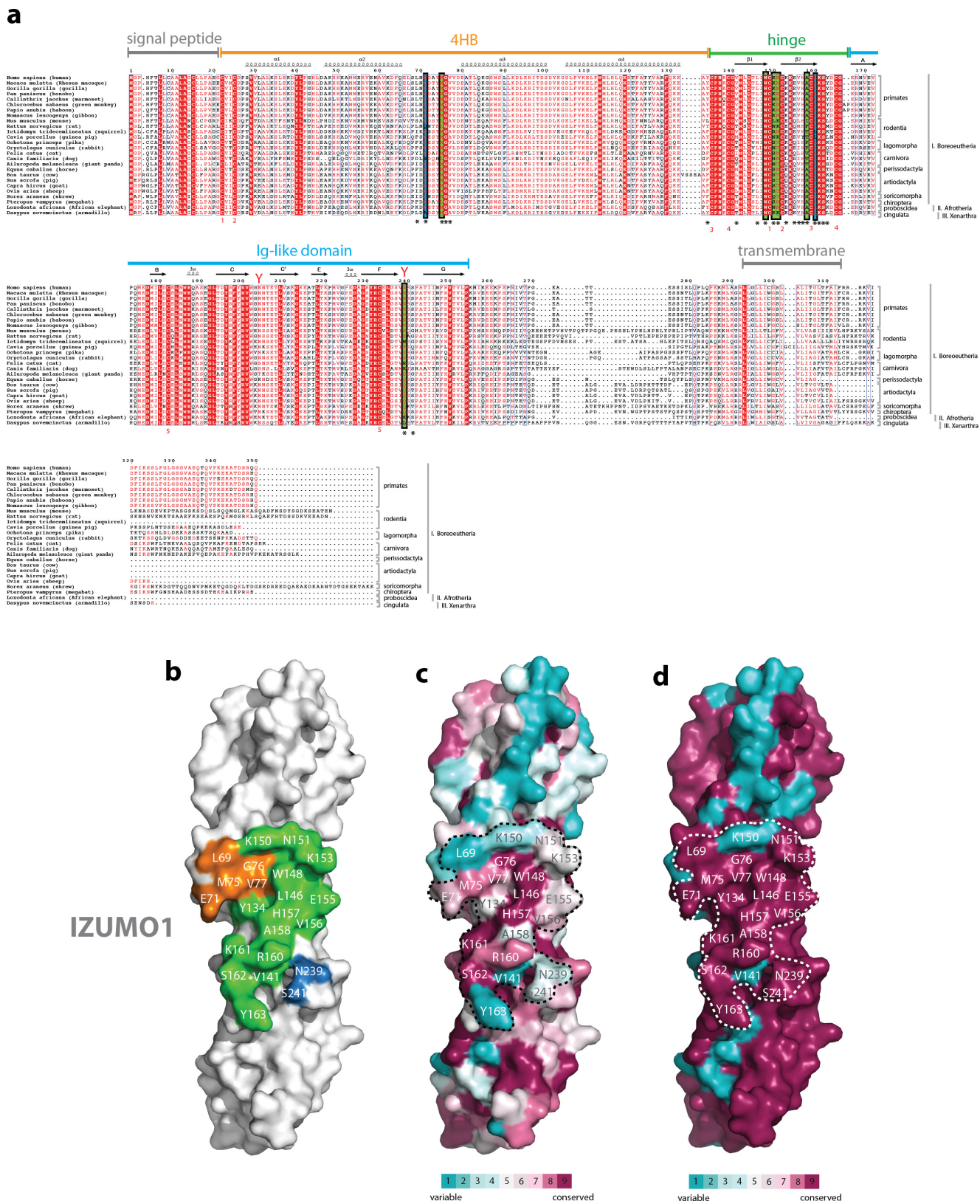
**Validation and structure analysis.** The stereochemical quality of all the refined models was validated using MolProbity<sup>45</sup>, PROCHECK<sup>46</sup> and Coot<sup>43</sup>. No residues were identified in disallowed regions of the Ramachandran plot. Moreover,

the  $R$  values,  $B$ -factors, and r.m.s.d. bond lengths and angles of all structural models are consistent with other deposited structures determined at similar resolutions, as validated by polygon.phenix<sup>35</sup>. All structural representations were prepared using PyMOL (v. 1.7.4 Schrödinger, LLC.)

**SAXS data collection and reconstruction.** SAXS provides medium-resolution visualization of global structural conformational changes in solution. SAXS experiments were performed by mail-in SAXS on beamline 12.3.1 (SIBYLS) at the Advanced Light Source. Various protein concentrations of tag-cleaved, deglycosylated wild-type and mutant IZUMO1<sub>22–254</sub>, JUNO<sub>20–228</sub> and IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complexes, along with matching buffer obtained from SEC (10 mM Tris-HCl (pH 8.0), 150 mM NaCl and 2% (v/v) glycerol), were loaded into a 96-well PCR plate (Axygen) and stored at 4 °C before data collection. Samples were loaded into the SAXS sample cell using a Hamilton syringe robot. For the wild-type IZUMO1<sub>22–254</sub>, JUNO<sub>20–228</sub> and IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex, data were collected at a wavelength ( $\lambda$ ) of 1.0 Å using a MarCCD 165 detector (Rayonix) positioned at a distance of 1.5 m, resulting in scattering vector  $q$  of  $0.01 \text{ \AA}^{-1} < q < 0.32 \text{ \AA}^{-1}$  (where  $q = 4\pi \sin(\theta/2)/\lambda$  and  $\theta$  is the scattering angle). Each data set was recorded at 283 K in a succession of three X-ray exposures of 0.5, 1, 2 and 5 s. For the mutant IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complexes, data were collected at 1.0 Å wavelength using a Pilatus3 × 2 M detector (Dectris) ( $0.01 \text{ \AA}^{-1} < q < 0.55 \text{ \AA}^{-1}$ ). Data for the mutant IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complexes were recorded in a time slicing mode of 0.5-s exposures over 15 s (30 frames per sample). Data for buffer blanks were collected before each protein image and subsequently buffer subtracted. Sample radiation damage was assessed by overlaying short and long exposures and detecting for any shifts in the scattering curves using the program SCATTER<sup>47</sup>. Concentration and aggregation effects were detected by comparing the lowest scattering angles for each of the protein samples. Fits to the Guinier region were made using autoRg. To maximize the signal-to-noise ratio, the SAXS scattering curve at the highest concentration that is free of interparticle interference was used for subsequent analysis. The characteristic real-space distance distribution function,  $P(r)$ , was determined from the scattering data using an indirect Fourier transformation and the maximum dimension,  $D_{\text{max}}$ <sup>48</sup>. All *ab initio* reconstructions of molecular envelopes from SAXS data were performed using the program DAMMIN<sup>49</sup>. Twenty-three DAMMIN models were superimposed and averaged by the program DAMAVER<sup>50</sup> to obtain a consensus averaged structure. Alignment of the SAXS reconstructions with the final refined crystal structures was performed using Chimera<sup>51</sup>.

**Deuterium exchange mass spectrometry (DXMS).** DXMS, which measures kinetics of backbone amide solvent exchange, provides local residue-level conformational dynamics. Prior to performing deuterium exchange experiments, the optimal proteolysis conditions were established as previously described<sup>52,53</sup> to maximize peptide sequence coverage of tag-cleaved, deglycosylated IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub>. Briefly, 1 µl of diluted protein stock solution (2 mg ml<sup>−1</sup> in 10 mM Tris (pH 7.2), 150 mM NaCl) was mixed with 5 µl quench buffer (6.4 M GuHCl and 1.0 M TCEP in 0.8% (v/v) formic acid, 16.6% (v/v) glycerol). After incubating on ice for various times (2, 5, 10, 15 and 30 min), the quenched samples were mixed with 24 µl dilution buffer (0.8% (v/v) formic acid, 16.6% (v/v) glycerol) and then subjected to proteolysis and LC-MS analysis. The IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex was formed by mixing IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> at 1:1.2 or 1.2:1 molar ratios and incubating the samples at 22 °C for 2 h. Deuterium exchange was initiated by mixing 3.5 µl of protein stock solution (IZUMO1<sub>22–254</sub>, IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub>, JUNO<sub>20–228</sub> or JUNO<sub>20–228</sub>–IZUMO1<sub>22–254</sub>) with 7 µl D<sub>2</sub>O buffer (8.3 mM Tris (pH 7.2), 150 mM NaCl in D<sub>2</sub>O,  $pD_{\text{read}}$  7.2) and incubating at 0 °C for 10, 100, 1000, 10,000 and 100,000 s. At indicated times, 2.1 µl of exchange samples were added to 10.5 µl quench solution to stop the D<sub>2</sub>O exchange reaction. After 5 min (IZUMO1<sub>22–254</sub> or IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub>) or 10 min (JUNO<sub>20–228</sub> or JUNO<sub>20–228</sub>–IZUMO1<sub>22–254</sub>) incubation on ice, quenched samples were diluted by addition of 48 µl of ice-cold dilution buffer, and then immediately frozen on dry ice and stored at −80 °C. The non-deuterated control samples and equilibrium-deuterated control samples were also prepared by mixing protein with H<sub>2</sub>O buffer (8.3 mM Tris (pH 7.2), 150 mM NaCl in H<sub>2</sub>O) and equilibrium-deuterated buffer (0.8% (v/v) formic acid in 99.9% D<sub>2</sub>O)<sup>54</sup>. The frozen samples were then thawed at 5 °C and passed over an immobilized pepsin column (16-µl bed volume) at a flow rate of 20 µl min<sup>−1</sup>. The resulting peptides were collected on a C<sub>18</sub> trap for desalting and separated by a Magic AQ C18 reverse phase column (Michrom BioResources) using a linear gradient of acetonitrile from 6.4% to 38.4% over 30 min. MS analysis was performed using the Orbitrap Elite Mass Spectrometer (ThermoFisher Scientific), with a capillary temperature of 200 °C. Data were acquired in both data-dependent MS/MS mode and MS1 profile mode, and the data were analysed by Proteome Discoverer software and DXMS Explorer<sup>55</sup> (Sierra Analytics Inc.).

29. Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668–W673 (2004).
30. Sultana, A. & Lee, J. E. Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. *Curr. Prot. Protein Sci.* **79**, 19.25.1–19.26 (2015).
31. D'Arcy, A., Villard, F. & Marsh, M. An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr. D* **63**, 550–554 (2007).
32. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
33. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
34. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
35. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
36. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **59**, 1131–1137 (2003).
37. Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N. & Adams, P. D. Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallogr. D* **64**, 99–107 (2008).
38. Dauter, Z. Twinned crystals and anomalous phasing. *Acta Crystallogr. D* **59**, 2004–2016 (2003).
39. Yeates, T. O. Simple statistics for intensity data from twinned specimens. *Acta Crystallogr. A* **44**, 142–144 (1988).
40. Fisher, R. G. & Sweet, R. M. Treatment of diffraction data from crystals twinned by merohedry. *Acta Crystallogr. A* **36**, 755–760 (1980).
41. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. D* **68**, 352–367 (2012).
42. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
43. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
44. Terwilliger, T. C. *et al.* Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr. D* **64**, 515–524 (2008).
45. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
46. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
47. Förster, S., Apostol, L. & Bras, W. Scatter : software for the analysis of nano- and mesoscale small-angle scattering. *J. Appl. Crystallogr.* **43**, 639–646 (2010).
48. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481 (2013).
49. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
50. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
51. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
52. Marsh, J. J. *et al.* Structural insights into fibrinogen dynamics using amide hydrogen/deuterium exchange mass spectrometry. *Biochemistry* **52**, 5491–5502 (2013).
53. Sheerin, D. J. *et al.* Inter- and intra-molecular interactions of *Arabidopsis thaliana* DELLA protein RGL1. *Biochem. J.* **435**, 629–639 (2011).
54. Li, S. *et al.* Mechanism of intracellular cAMP sensor Epac2 activation: cAMP-induced conformational changes identified by amide hydrogen/deuterium exchange mass spectrometry (DXMS). *J. Biol. Chem.* **286**, 17889–17897 (2011).
55. Zhang, Z. & Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**, 522–531 (1993).

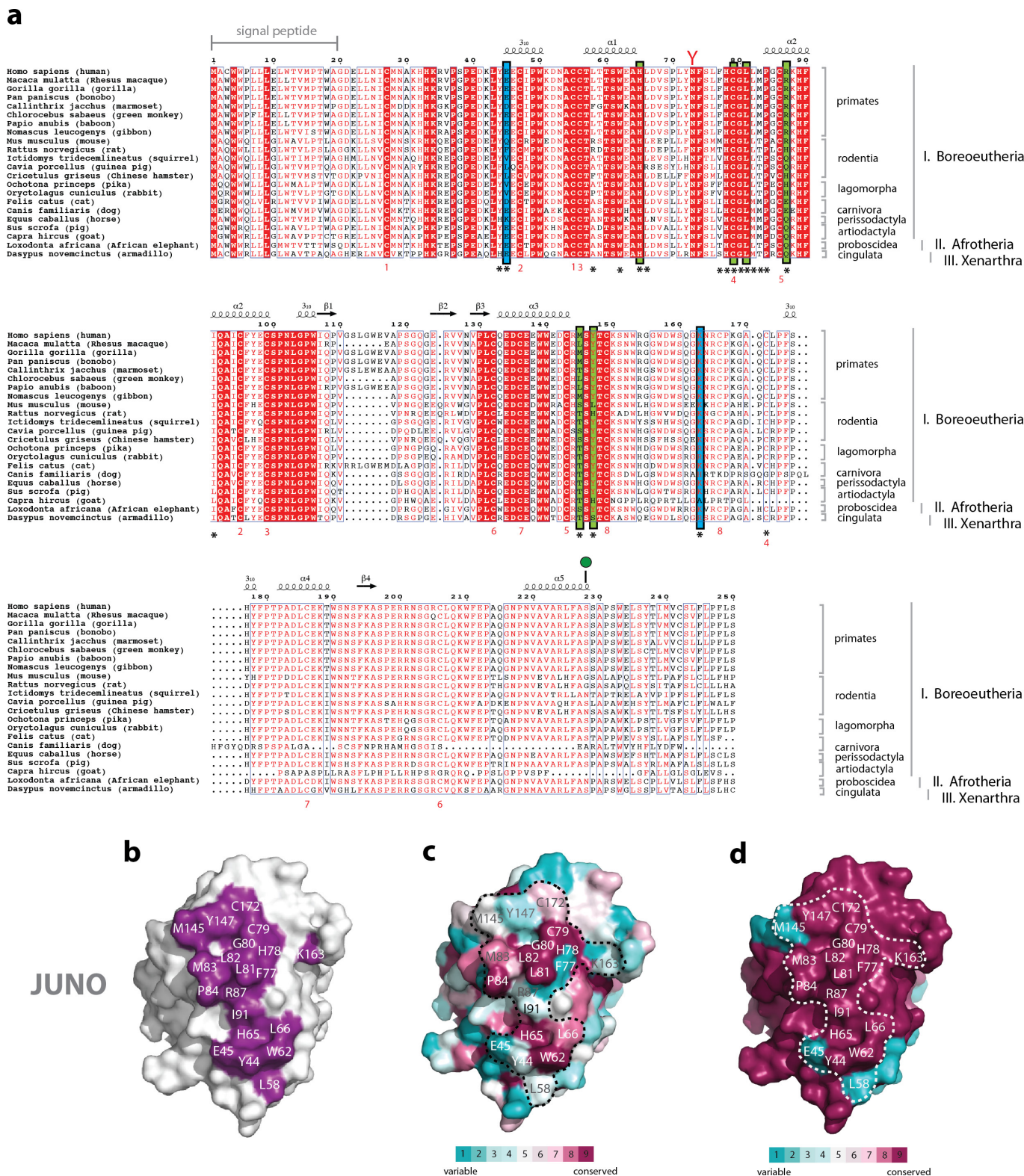


**Extended Data Figure 1** | See next page for caption.

**Extended Data Figure 1 | Conservation of IZUMO1 residues.**

**a**, Alignment of IZUMO1 protein sequences from various mammals. IZUMO1 sequences from *Homo sapiens* (human; GenBank: BAD91012.1), *Macaca mulatta* (rhesus macaque; GenBank: EHH30233.1), *Gorilla gorilla* (gorilla; Uniprot: G3QFY5), *Pan paniscus* (bonobo; NCBI: XP\_003814124.1), *Callithrix jacchus* (marmoset; Uniprot: F7H859), *Chlorocebus sabaeus* (green monkey; Uniprot: A0A0D9S2Z4), *Papio anubis* (baboon; Uniprot: A0A0A0MU86), *Nomascus leucogenys* (gibbon; Uniprot: G1QXF7), *Mus musculus* (mouse; GenBank: BAD91011.1), *Rattus norvegicus* (rat; GenBank: BAD91013.1), *Ictidomys tridecemlineatus* (squirrel; Uniprot: I3N2L9), *Cavia porcellus* (guinea pig; Uniprot: H0UTJ7), *Ochotona princeps* (pika; NCBI: XP\_004597241.1), *Oryctolagus cuniculus* (rabbit; Uniprot: G1TVX5), *Felis catus* (cat; NCBI: XP\_006941089.1), *Canis familiaris* (dog; Uniprot: F6UM65), *Ailuropoda melanoleuca* (giant panda, Uniprot: G1M882), *Equus caballus* (horse; Uniprot: F6YE25), *Bos taurus* (cow; Uniprot: E1BDA8), *Sus scrofa* (pig; Uniprot: F1RIQ7), *Capra hircus* (goat; Uniprot: C6ZEA2),

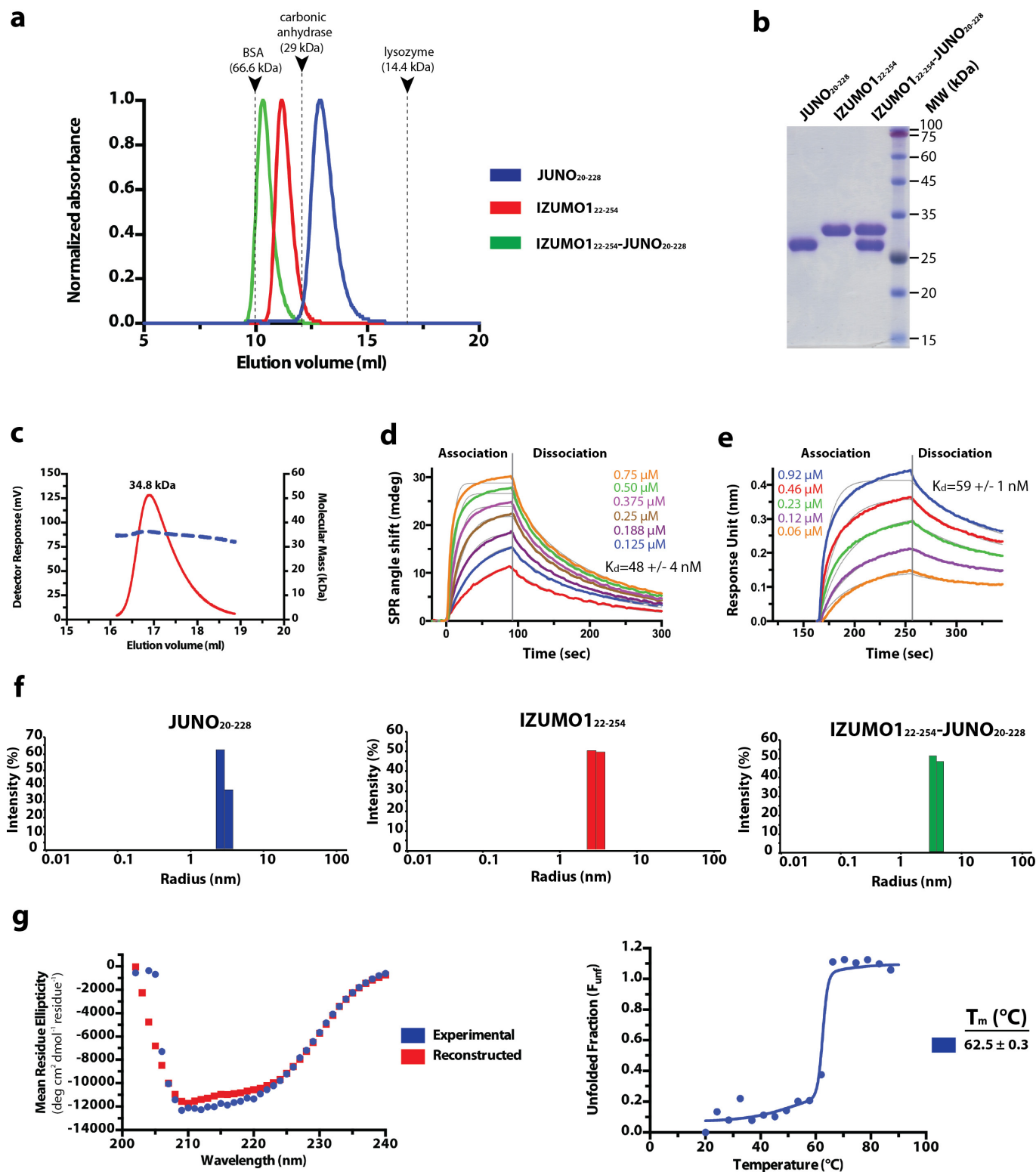
*Ovis aries* (sheep; Uniprot: W5PRD0), *Sorex araneus* (shrew; NCBI: XP\_004619786.1), *Pteropus vampyrus* (megabat; NCBI: XP\_011372928.1), *Loxodonta africana* (African elephant; NCBI: XP\_003406572.1), and *Dasyus novemcinctus* (armadillo; NCBI: XP\_004451154.1) are aligned. Red boxes indicate complete conservation of a given amino acid. N-linked glycosylation sequons (N-X-S/T) are indicated by red-coloured Y-shaped symbols. Secondary structural elements observed in the crystal structure of IZUMO1 are shown as arrows for  $\beta$ -strands and coils for  $\alpha$ -helices. Residues that interact with JUNO are identified with asterisks, with those that form salt bridges and hydrogen bonds highlighted in blue and green boxes, respectively. Cysteine pairs involved in disulfide bond formation are numbered in red underneath each sequence. **b**, Footprint of JUNO on the molecular surface of IZUMO1. **c**, **d**, Representation of surface residue conservation, calculated using ConSurf and the alignment of all mammalian IZUMO1 (**c**) or primate-only IZUMO1 (**d**) sequences from Extended Data Fig. 1a. Degree of residue conservation is coloured in a gradient from high (burgundy) to low (cyan) variability.



Extended Data Figure 2 | See next page for caption.

**Extended Data Figure 2 | Conservation of JUNO residues.** **a**, Alignment of JUNO protein sequences from various mammals. JUNO/FOLR- $\delta$  sequences from *H. sapiens* (human; NCBI: NP\_001186135.1), *M. mulatta* (rhesus macaque; NCBI: NP\_001180734.1), *G. gorilla* (gorilla; NCBI: XP\_004052029.1), *P. paniscus* (bonobo; NCBI: XP\_003813838.1), *C. jacchus* (marmoset; NCBI: XP\_009005477.1), *C. sabaues* (green monkey; Uniprot: A0A0D9S1B0), *P. anubis* (baboon; NCBI: XP\_009185381.1), *N. leucogenys* (gibbon; Uniprot: G1R639), *M. musculus* (mouse; NCBI: NP\_075026.1), *R. norvegicus* (rat; NCBI: XP\_001072998.2), *I. tridecemlineatus* (squirrel; NCBI: XP\_005337246.1), *C. porcellus* (guinea pig; NCBI: XP\_003468609.1), *Cricetulus griseus* (Chinese hamster; NCBI: XP\_003506544.1), *O. princeps* (pika; NCBI: XP\_012782378.1), *O. cuniculus* (rabbit; Uniprot: G1T5D7), *F. catus* (cat; NCBI: XP\_011284828.1), *C. familiaris* (dog; Uniprot: E2RTK1), *E. caballus* (horse; NCBI: XP\_001491306.1), *S. scrofa* (pig; Uniprot: F1STK4), *C. hircus* (goat; NCBI: XP\_013824827.1), *L. africana* (African elephant; NCBI: XP\_010593777.1), and *D. novemcinctus* (armadillo; NCBI:

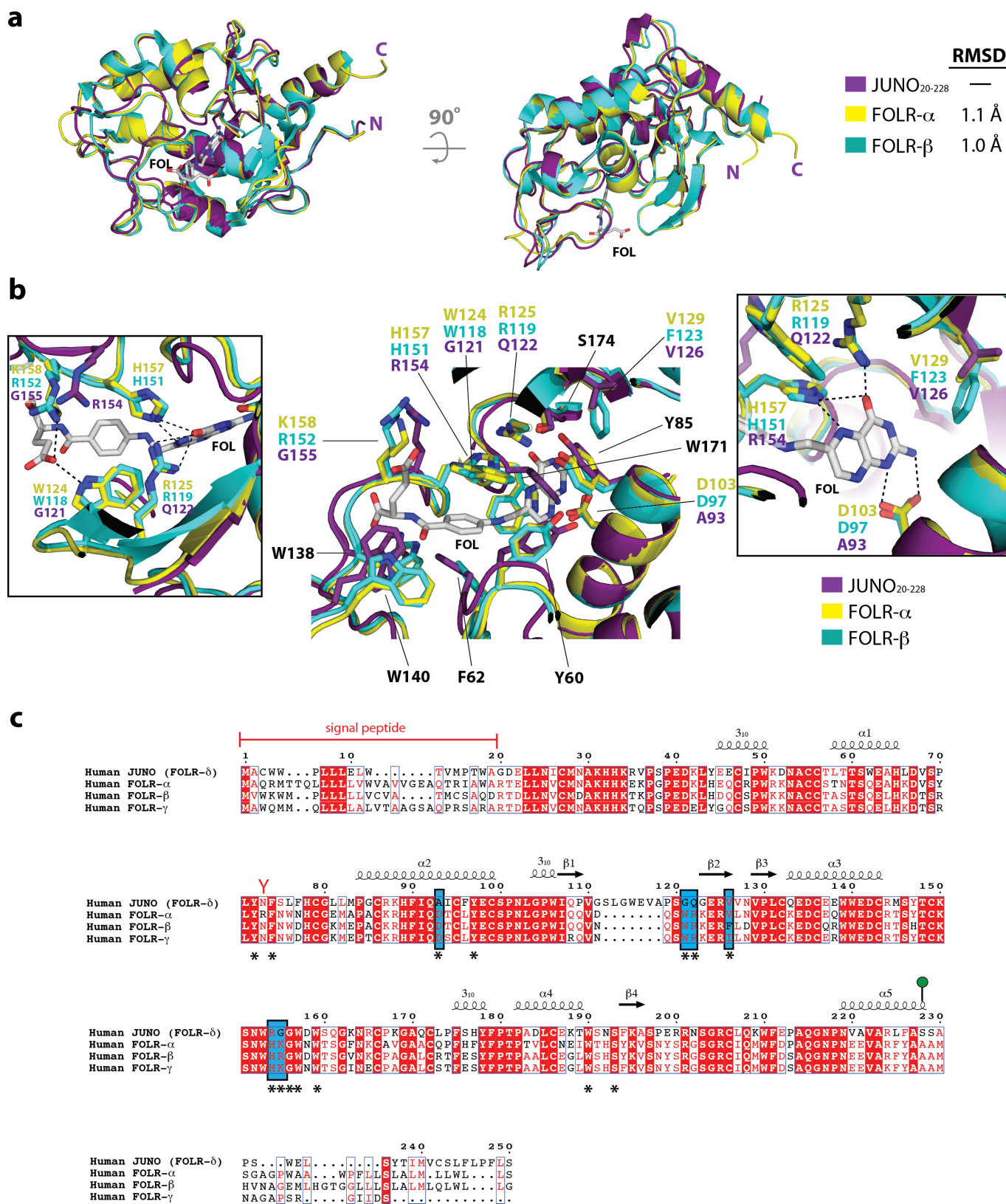
XP\_004471965.1) are aligned. Red boxes indicate complete conservation of a given amino acid. N-linked glycosylation sequons (N-X-S/T) are indicated by red-coloured Y-shaped symbols. JUNO is anchored to the plasma membrane through a GPI anchor at Ser228 (shown as a green lollipop). Secondary structural elements observed in the crystal structure of JUNO are shown as arrows for  $\beta$ -strands and coils for  $\alpha$ -helices. Residues that interact with IZUMO1 are identified with asterisks underneath the sequence, with those that form salt bridges and hydrogen bonds highlighted in blue and green boxes, respectively. Cysteine pairs involved in disulfide bond formation are numbered in red underneath each sequence. **b**, Footprint of IZUMO1 on the molecular surface of JUNO. **c**, **d**, Representation of surface residue conservation, calculated using ConSurf and the alignment of all mammalian JUNO (**c**) or primate-only JUNO sequences (**d**) from Extended Data Fig. 2a. Degree of residue conservation is coloured in a gradient from high (burgundy) to low (cyan) variability.



Extended Data Figure 3 | See next page for caption.

**Extended Data Figure 3 | Purification and characterization of IZUMO1 and JUNO.** **a**, Superdex-75 10/300 GL size-exclusion chromatograms of JUNO<sub>20–228</sub>, IZUMO1<sub>22–254</sub>, and the IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex. Eluted peak positions of protein standards are marked with triangles and dashed lines. **b**, Coomassie-stained SDS–PAGE analysis of the purified IZUMO1<sub>22–254</sub>, JUNO<sub>20–228</sub> and IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex. For gel source data, see Supplementary Fig. 1c. **c**, Size-exclusion chromatography with inline multi-angle light scattering (SEC–MALS) profile of glycosylated human IZUMO1<sub>22–268</sub>. The detector response unit (mV) and molecular mass (kDa) are plotted against the elution volume from a Superdex-200 Increase 10/300 GL size exclusion column. SEC–MALS reveals an apparent molecular mass of 34.8 kDa (dashed blue line), which corresponds to a monomeric species. **d**, Surface plasmon resonance (SPR) binding affinity and kinetic analysis of the human IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub> interaction. Human JUNO<sub>20–228</sub> was amine-coupled to the SPR sensor chip. Kinetic parameters were derived from a Langmuir 1:1 binding model. **e**, Biolayer interferometry (BLI) kinetic analysis of the interaction between human IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub>.

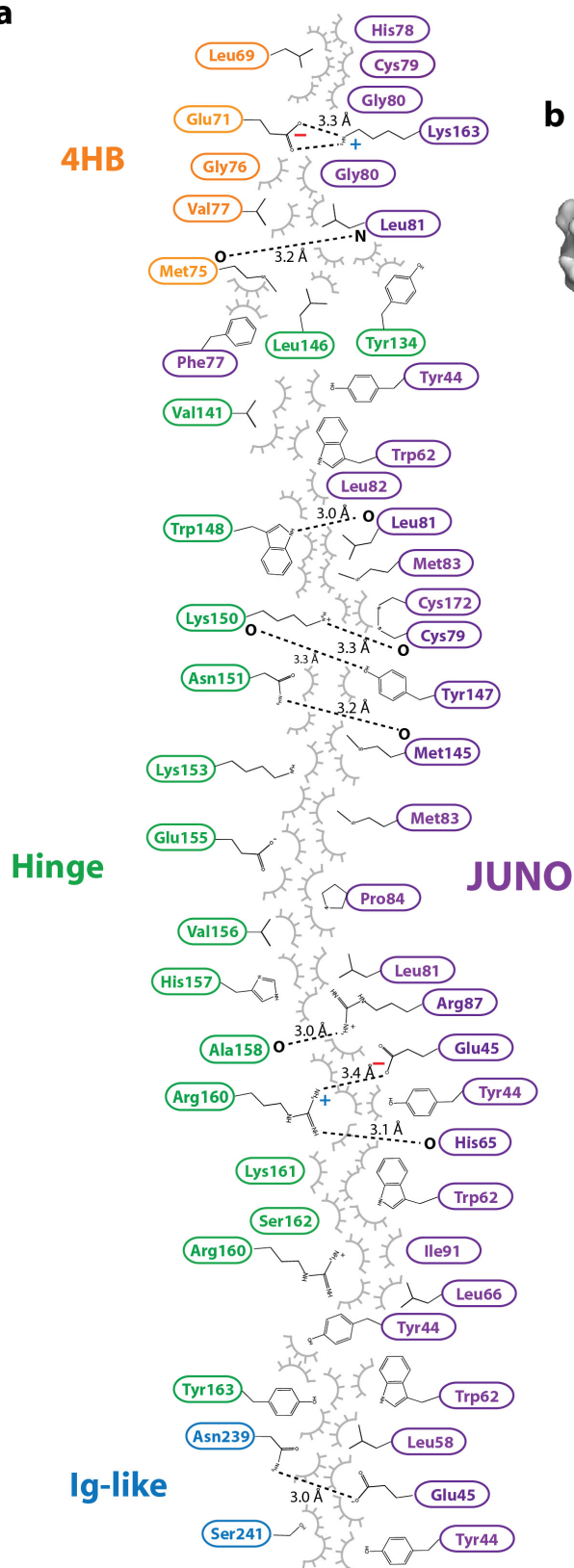
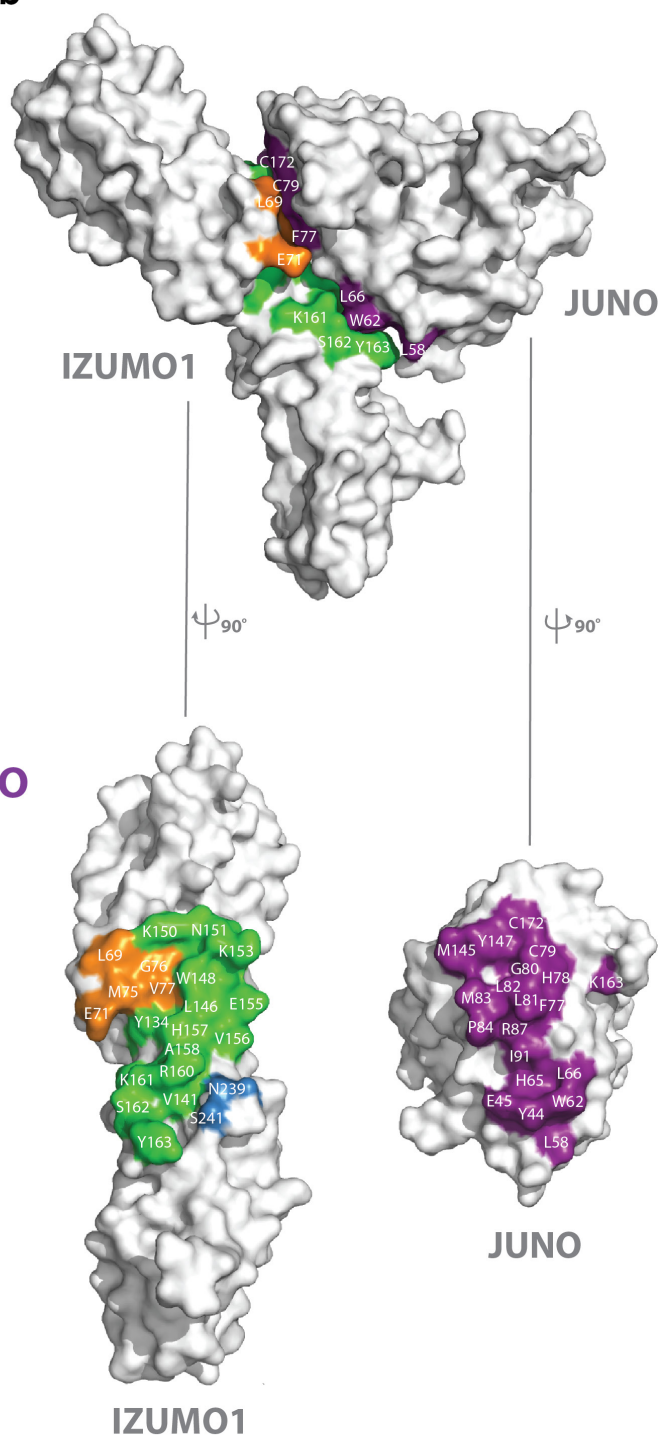
Human JUNO<sub>20–228</sub> was biotinylated and coupled to streptavidin-coated biosensors. Kinetic parameters were derived from a 1:1 binding model. The experimental curves are shown in colour superimposed with the fitted curves indicated as grey lines. **f**, A size distribution histogram from dynamic light scattering (DLS) measurements of IZUMO1<sub>22–254</sub>, JUNO<sub>20–228</sub> and IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex at 5 mg ml<sup>−1</sup>. IZUMO1<sub>22–254</sub>, JUNO<sub>20–228</sub> and IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> display hydrodynamic radii ( $R_H$ ) of ~3.0 nm, ~2.9 nm and ~3.9 nm, respectively. **g**, Circular dichroism (CD) wavelength scan of human IZUMO1<sub>22–268</sub> (blue) at 25 °C shows mixed secondary structural characteristics. The crystal structure of IZUMO1<sub>22–268</sub> aligns well with the secondary structural content calculated from the CD spectrum (35%  $\alpha$ -helical, 24%  $\beta$ -strand and 41% random coil). A reconstructed CD wavelength scan (red) illustrates the agreement of the fit used in secondary structural content analysis. A CD thermal denaturation profile of human IZUMO1<sub>22–268</sub> at 222 nm is shown. The CD signal was normalized between 0 (folded) and 1 (unfolded), and plotted as a function of temperature. The  $T_m$  value indicates the midpoint of the melting transition.



Extended Data Figure 4 | See next page for caption.

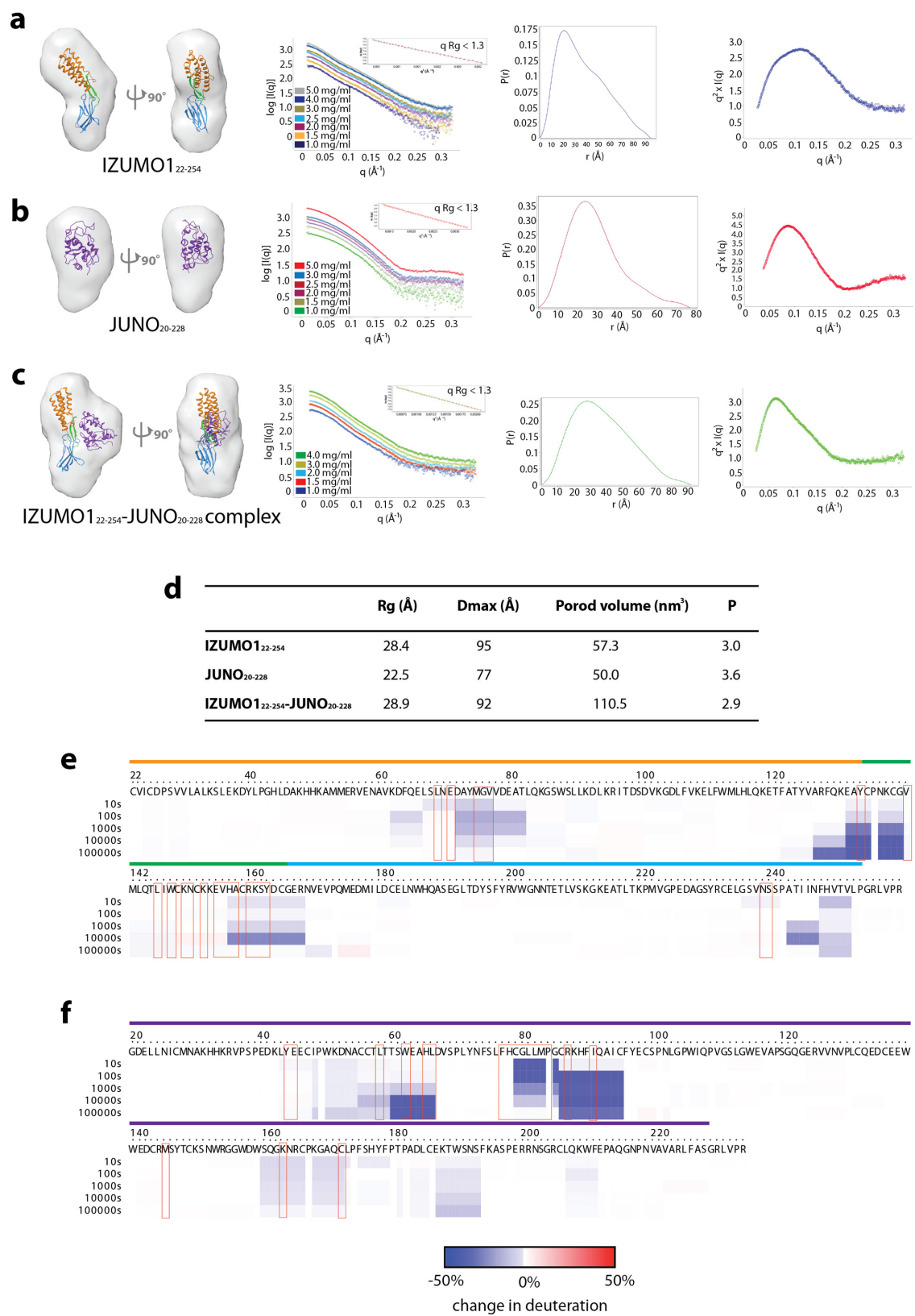
**Extended Data Figure 4 | Structural comparison of JUNO and the folate receptor family of proteins.** **a**, Structural superimposition of JUNO<sub>20–228</sub> with FOLR- $\alpha$  (PDB ID: 4LRH) and FOLR- $\beta$  (PDB ID: 4KMZ). Experimentally bound folate (FOL), shown in white sticks, from the FOLR- $\alpha$  structure is positioned in the active site. **b**, Superimposition of residues in the folate-binding site of human FOLR- $\alpha$  and FOLR- $\beta$ , and equivalent residues in human JUNO. Residue names shown in black are conserved among JUNO, FOLR- $\alpha$  and FOLR- $\beta$ , and are numbered on the basis of the FOLR- $\alpha$  sequence. Inset boxes highlight the residue differences between JUNO, FOLR- $\alpha$  and FOLR- $\beta$ . Key hydrogen bond interactions are shown as dashed black lines. Mutagenesis studies showed that replacement of D103 or D97 in FOLR- $\alpha$  or FOLR- $\beta$ , respectively, which form strong interactions to the N1 and N2 nitrogen atoms of the pterin moiety, results in a decrease in affinity of more than one order of magnitude<sup>15</sup>. Six folate-binding residues observed in FOLR- $\alpha$  and FOLR- $\beta$  (FOLR- $\alpha$ /FOLR- $\beta$ : D103/D97, W124/W118, R125/R119, V129/F123,

H157/H151, and K158/R152) are not conserved in JUNO. Four of these residues (FOLR- $\alpha$ /FOLR- $\beta$ : D103/D97, W124/W118, R125/R119, and H157/H151) form key hydrogen bonds to anchor folate in the active site. In JUNO, the substituted residues are not able to maintain the extensive hydrogen bond network seen in FOLR- $\alpha$  and FOLR- $\beta$  to folate. **c**, *H. sapiens* FOLR- $\alpha$  (Uniprot: P15328), FOLR- $\beta$  (Uniprot: P14207), FOLR- $\gamma$  (Uniprot: P41439) and FOLR- $\delta$  (Uniprot: A6ND01) are aligned. Red boxes indicate complete conservation of a given amino acid. N-linked glycosylation sequons (N-X-S/T) are indicated by red-coloured Y-shaped symbols. JUNO is anchored to the plasma membrane through a GPI anchor at Ser228 (shown as a green lollipop). Experimentally determined secondary structural elements are shown as arrows for  $\beta$ -strands and coils for  $\alpha$ -helices. Key folate-binding residues, identified from the FOLR- $\alpha$  and FOLR- $\beta$  crystal structures, are identified with an asterisk underneath the sequence. Key residue differences between JUNO, FOLR- $\alpha$  and FOLR- $\beta$  folate binding sites are highlighted in a blue box.

**a****b**

**Extended Data Figure 5 | IZUMO1–JUNO interface.** **a**, 2D schematic of the interactions between IZUMO1<sub>22–254</sub> and JUNO<sub>20–228</sub>. Residues from the IZUMO1 4HB, hinge, and Ig-like regions and from JUNO are coloured orange, green, blue and purple, respectively. Hydrogen-bond interactions are shown as dashed lines, and van der Waals forces are depicted as grey semi-circles. **b**, Footprints of JUNO on the surface of IZUMO1 and of

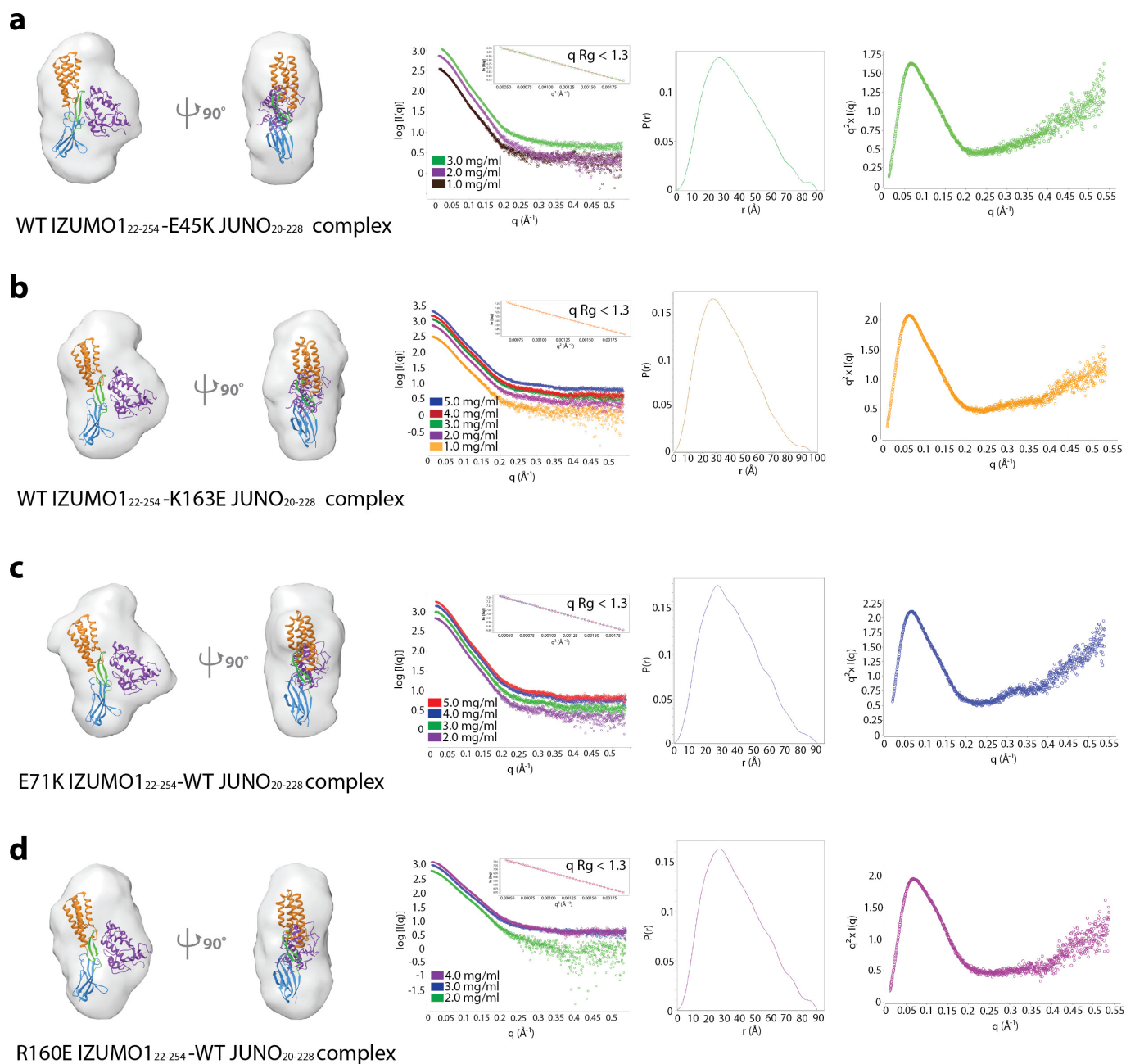
IZUMO1 on the surface of JUNO. The molecular surfaces of IZUMO1 and JUNO are coloured white with residues forming interactions coloured as in **a**. No N-linked glycans on either IZUMO1<sub>22–254</sub> or JUNO<sub>20–228</sub> are involved in binding. Formation of this interface results in a calculated free energy gain of  $-10.4 \text{ kcal mol}^{-1}$ .



Extended Data Figure 6 | See next page for caption.

**Extended Data Figure 6 | Hybrid structural analysis of human IZUMO1 and JUNO in a solution state.** **a–c**, *Ab initio* SAXS reconstruction, experimental scattering curves, normalized pair distance distribution function,  $P(r)$  and Kratky plot showing the degree of flexibility of IZUMO1<sub>22–254</sub> (**a**), JUNO<sub>20–228</sub> (**b**), and the IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex (**c**). No concentration-dependent or radiation effects were observed in the SAXS data. The inset box in the experimental scattering data shows linearity in the Guinier plot at low  $q$  ( $qR_g < 1.3$ ). The IZUMO1<sub>22–254</sub>, JUNO<sub>20–228</sub> and IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub> complex crystal structures were docked into the SAXS reconstructed molecular envelopes. The boomerang shape and upright conformation seen in the

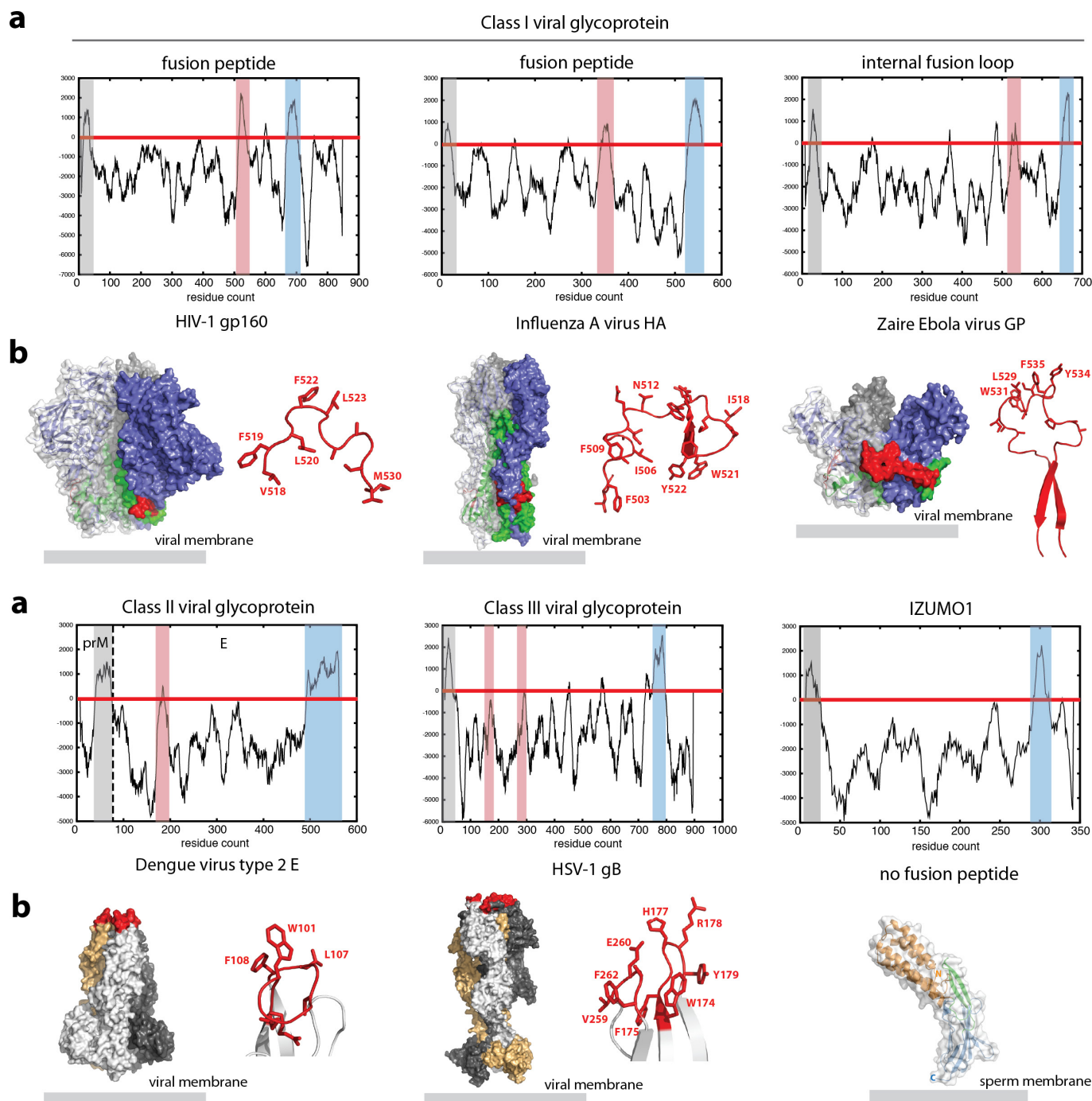
crystal structures of unbound and bound IZUMO1<sub>22–254</sub>, respectively, were recapitulated by the SAXS reconstructions. **d**, Summary of the experimentally derived SAXS parameters for IZUMO1<sub>22–254</sub>, JUNO<sub>20–228</sub> and IZUMO1<sub>22–254</sub>–JUNO<sub>20–228</sub>. The program SCATTER<sup>47</sup> was used to calculate the radius of gyration ( $R_g$ ) and maximum linear dimension ( $D_{max}$ ), and to perform Porod–Debye analysis to obtain the Porod volume and  $P$  coefficient. **e**, **f**, Comparative deuterium exchange mass spectrometry (DXMS) profile of unbound and bound IZUMO1<sub>22–254</sub> (**e**) and JUNO<sub>20–228</sub> (**f**). The plots reveal the change in individual deuterium exchange for all observable residues. The coloured lines above the residue numbers correspond to the observed regions in the crystal structures.

**e**

	Rg (Å)	Dmax (Å)	Porod volume (nm <sup>3</sup> )	P
E71K IZUMO1 <sub>22-254</sub> -WT JUNO <sub>20-228</sub>	29.5	91	109.4	2.9
R160E IZUMO1 <sub>22-254</sub> -WT JUNO <sub>20-228</sub>	29.5	90	110.7	2.8
WT IZUMO1 <sub>22-254</sub> -E45K JUNO <sub>20-228</sub>	29.5	90	110.5	2.8
WT IZUMO1 <sub>22-254</sub> -K163E JUNO <sub>20-228</sub>	29.6	97	110.6	3.0

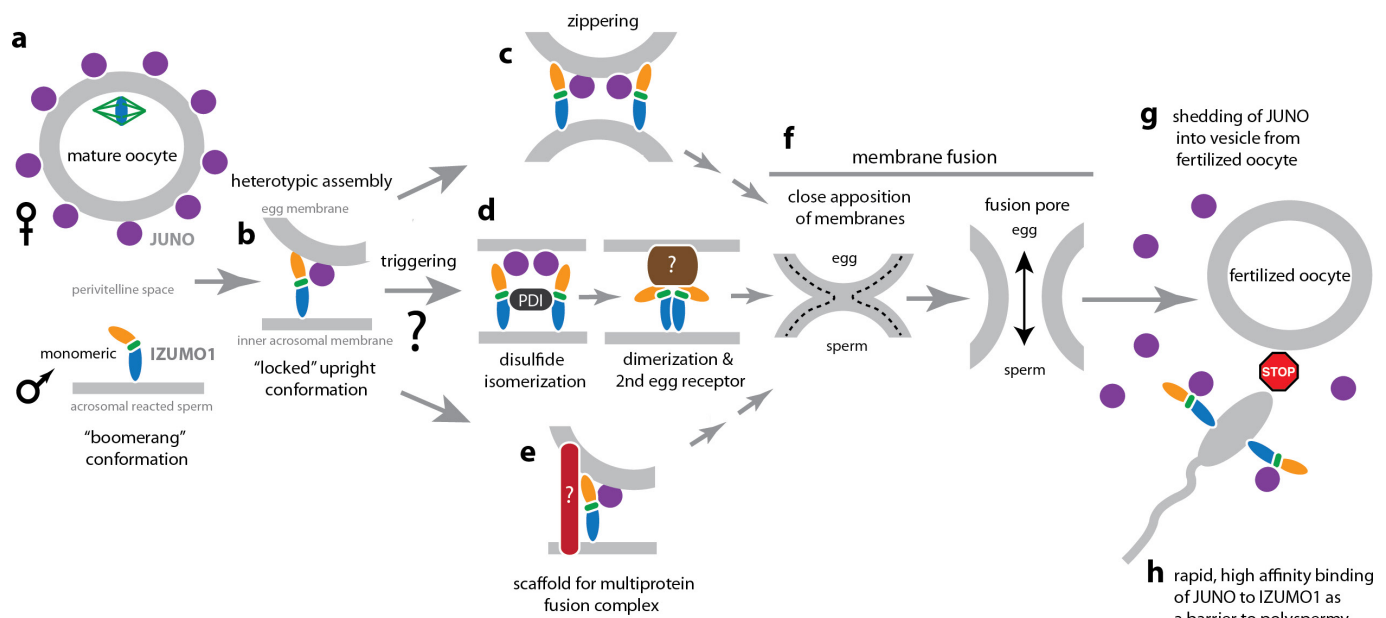
**Extended Data Figure 7 | SAXS reconstruction of IZUMO1 and JUNO mutants.** **a–d**, *Ab initio* SAXS reconstruction, experimental scattering curves, normalized pair distance distribution function,  $P(r)$  and Kratky plot showing the degree of flexibility of wild-type (WT) IZUMO1<sub>22-254</sub>-JUNO<sub>20-228</sub> (E45K) (**a**), IZUMO1<sub>22-254</sub>(WT)-JUNO<sub>20-228</sub>(K163E) (**b**), IZUMO1<sub>22-254</sub>(E71K)-JUNO<sub>20-228</sub>(WT) (**c**), and IZUMO1<sub>22-254</sub>(R160E)-JUNO<sub>20-228</sub>(WT) (**d**) complexes. No concentration-dependent or radiation effects were observed. The inset box shows linearity in the Guinier plot

at low  $q$  ( $qRg < 1.3$ ). The IZUMO1<sub>22-254</sub>(WT)-JUNO<sub>20-228</sub>(WT) complex crystal structure was docked into the SAXS reconstructed molecular envelopes. **e**, Summary of the experimentally derived SAXS parameters for the various IZUMO1-JUNO complexes. The program SCATTER<sup>47</sup> was used to calculate the radius of gyration ( $R_g$ ) and maximum linear dimension ( $D_{max}$ ), and to perform Porod-Debye analysis to obtain the Porod volume and  $P$  coefficient.



**Extended Data Figure 8 | Comparison of IZUMO1 with selected viral fusogens.** A common feature of many viral fusogens is the presence of a hydrophobic fusion peptide or fusion loop. **a**, Kyte and Doolittle hydropathy plots were calculated for IZUMO1, HIV-1 gp160, influenza A virus HA, Ebola virus glycoprotein (GP), Dengue virus type 2 E, and herpes simplex virus-1 gB to detect the presence of hydrophobic regions. Class I and class II viral fusion glycoproteins contain three clear hydrophobic regions corresponding to the signal peptide (grey), fusion peptide or loop (red) and the transmembrane anchor (blue). For class III viral glycoproteins, the presence of a signal peptide and transmembrane anchor are clear, but the hydrophobic fusion loop is formed by two discontinuous regions. This results in a lower hydropathy scale that is more difficult to detect. Two regions of hydrophobic residues cluster at the

tip of the glycoprotein (shown in red) and are thought to be the internal fusion loop. In all class I, II and III viral fusion glycoproteins, clustering of aromatic and hydrophobic residues in a loop or helical region is a hallmark feature of fusion proteins. In contrast, IZUMO1 clearly does not have any hydrophobic regions or structural features similar to the viral fusogens that could insert into the egg membrane. **b**, Molecular surface representation of class I, II, and III viral glycoproteins and IZUMO1. The fusion peptide or loop is shown as red sticks and also coloured red on the glycoprotein surface. For the class I viral glycoproteins, the metastable prefusion trimer is shown, with the receptor binding and fusion subunits shown in blue and green, respectively. For the class II and class III viral glycoproteins, the postfusion trimer is shown with three hydrophobic fusion loops clustered at the tip of the molecule.



**Extended Data Figure 9 | Model of IZUMO1 and JUNO in sperm-egg fertilization.** During fertilization, mature sperm undergoes an acrosome reaction and penetrates through the egg zona pellucida to reach the perivitelline space. The acrosome reaction also causes relocation of IZUMO1 to the sperm equatorial segment. **a**, IZUMO1 adopts a monomeric boomerang conformation on the surface of the sperm membrane. **b**, Upon binding to the JUNO egg receptor, IZUMO1 undergoes a conformational change. The 4HB region migrates towards the egg membrane. Moreover, the hinge region of IZUMO1 becomes more rigid and 'locks' the molecule into an upright position. The formation of the IZUMO1 and JUNO complex provides a direct physical link between the egg and sperm membranes. It is currently not clear whether IZUMO1 requires a post-JUNO binding event to trigger the fusion process, but at least three potential mechanisms are possible. **c**, The heterotypic assembly of IZUMO1 and JUNO, or a secondary conformational change in IZUMO1, may bring the egg and sperm membranes into close proximity for fusion to take place. **d**, Inoue *et al.* proposed that subsequent to IZUMO1-JUNO binding, a protein disulfide isomerase

(PDI) catalyses a thio-disulfide exchange reaction that leads to structural conformation change and dimerization of IZUMO1 (ref. 26). The IZUMO1 dimer releases JUNO and contacts a yet-to-be-discovered oocyte receptor that facilitates membrane fusion. **e**, Alternatively, IZUMO1 may act as a scaffold to recruit other sperm or egg protein partners to form a multiprotein fusion complex in a manner similar to some viral fusogens. **f**, The merger of the egg and sperm membranes will require the apposition of the two bilayers to initiate initial mixing of the outer membrane leaflets and formation of a hemifusion stalk. The hemifused bilayers open to form the full fusion pore. **g**, Following fusion, JUNO is rapidly shed into extracellular vesicles from the fertilized oocyte. Within 30–40 min, JUNO is weakly or barely detectable on the membrane surface of zona-intact or anaphase II-stage zona-free fertilized oocytes, and undetectable at the pronuclear stage<sup>6</sup>. **h**, IZUMO1 binds JUNO tightly and rapidly (BLI:  $K_d = 59 \pm 1$  nM,  $k_a = 1.15 \times 10^5$  M<sup>-1</sup> s<sup>-1</sup>; SPR:  $48 \pm 4$  nM,  $k_a = 4.2 \times 10^5$  M<sup>-1</sup> s<sup>-1</sup>), and once shed, JUNO is able to bind exposed IZUMO1 on incoming acrosomal-reacted sperm in the perivitelline space to act as a 'sperm-sink' to block polyspermy.

Extended Data Table 1 | IZUMO1–JUNO binding interface mutations

Technique	Protein	Role of residue	$K_d$	$k_a$ ( $\times 10^5$ ) ( $M\ s^{-1}$ )	$k_d$ ( $\times 10^{-3}$ ) ( $s^{-1}$ )
SPR	Wild-type	n/a	48 +/- 4 nM	4.2 +/- 0.2	19.4 +/- 0.9
BLI	Wild-type	n/a	59 +/- 1 nM	1.15 +/- 0.02	6.7 +/- 0.1
<b>Izumo1 mutants</b>					
BLI	E71A IZUMO1 <sub>22-254</sub>	salt bridge with JUNO K163; L2 leash	96 +/- 13 nM	1.43 +/- 0.04	13.7 +/- 0.2
BLI	E71K IZUMO1 <sub>22-254</sub>	salt bridge with JUNO K163; L2 leash	119 +/- 19 nM	1.65 +/- 0.04	19.5 +/- 0.3
BLI	D72A IZUMO1 <sub>22-254</sub>	H-bond to Q130 in unbound state	71 +/- 8 nM	1.78 +/- 0.04	12.6 +/- 0.2
BLI	Q130A IZUMO1 <sub>22-254</sub>	H-bond to D72 in unbound state	56 +/- 19 nM	1.27 +/- 0.03	6.8 +/- 0.1
BLI	W148A IZUMO1 <sub>22-254</sub>	conserved interface residue	<i>no binding detected</i>		
BLI	H157A IZUMO1 <sub>22-254</sub>	conserved interface residue	1.8 +/- 1 $\mu$ M	0.51 +/- 0.05	79 +/- 2
BLI	R160A IZUMO1 <sub>22-254</sub>	salt bridge with JUNO E45	730 +/- 78 nM	0.78 +/- 0.04	56.3 +/- 0.9
BLI	R160E IZUMO1 <sub>22-254</sub>	salt bridge with JUNO E45	2.2 +/- 1.2 $\mu$ M	0.32 +/- 0.02	51.2 +/- 1.1
<b>Juno mutants</b>					
BLI	E45A JUNO <sub>20-228</sub>	salt bridge with IZUMO1 R160	681 +/- 141 nM	0.37 +/- 0.01	24.7 +/- 0.5
BLI	E45K JUNO <sub>20-228</sub>	salt bridge with IZUMO1 R160	2.8 +/- 0.2 $\mu$ M	0.23 +/- 0.02	62.4 +/- 1.7
BLI	W62A JUNO <sub>20-228</sub>	conserved interface residue	361 +/- 24 nM	1.14 +/- 0.04	41.2 +/- 0.6
BLI	L81A JUNO <sub>20-228</sub>	conserved interface residue	9.1 +/- 0.4 $\mu$ M	0.4 +/- 0.3	349 +/- 23
BLI	K163A JUNO <sub>20-228</sub>	salt bridge with IZUMO1 E71	97 +/- 9 nM	1.38 +/- 0.03	13.3 +/- 0.2
BLI	K163E JUNO <sub>20-228</sub>	salt bridge with IZUMO1 E71	134 +/- 4 nM	1.81 +/- 0.05	24.2 +/- 0.3

\*kinetic and binding affinity values are presented as the mean of technical triplicates ( $n=3$ ) and errors as the standard deviation of the mean

$K_d$ - dissociation constant

$k_a$ - rate of association

$k_d$ - rate of dissociation

All experiments were performed with technical triplicates ( $n=3$ ), with mean  $K_d$  values  $\pm$  s.e.m.

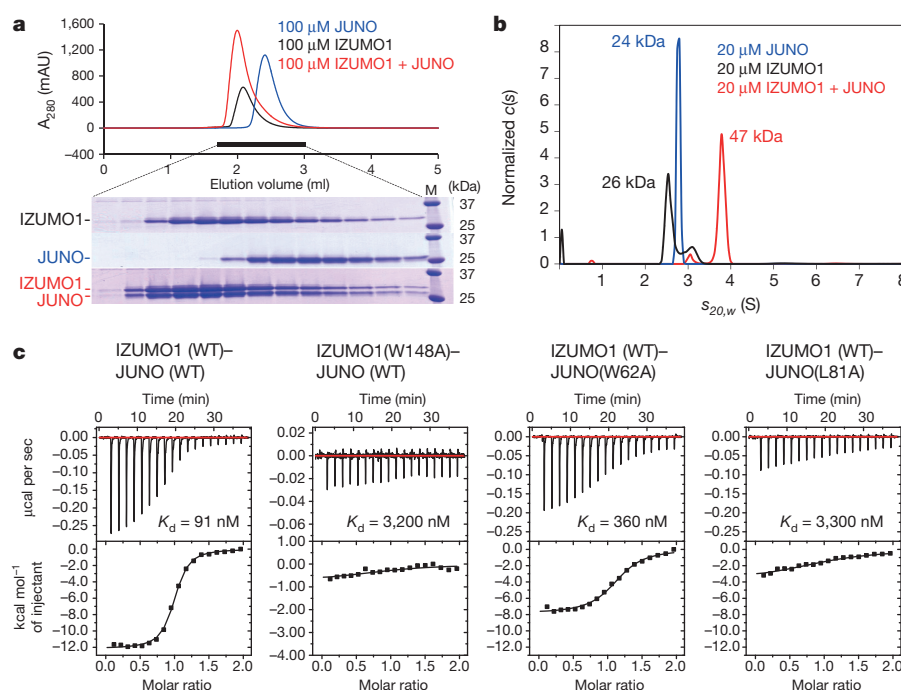
# Structure of IZUMO1–JUNO reveals sperm–oocyte recognition during mammalian fertilization

Umeharu Ohto<sup>1\*</sup>, Hanako Ishida<sup>1\*</sup>, Elena Krayukhina<sup>2</sup>, Susumu Uchiyama<sup>2,3</sup>, Naokazu Inoue<sup>4</sup> & Toshiyuki Shimizu<sup>1</sup>

Fertilization is a fundamental process in sexual reproduction, creating a new individual through the combination of male and female gametes<sup>1–4</sup>. The IZUMO1 sperm membrane protein<sup>5</sup> and its counterpart oocyte receptor JUNO<sup>6</sup> have been identified as essential factors for sperm–oocyte interaction and fusion. However, the mechanism underlying their specific recognition remains poorly defined. Here, we show the crystal structures of human IZUMO1, JUNO and the IZUMO1–JUNO complex, establishing the structural basis for the IZUMO1–JUNO-mediated sperm–oocyte interaction. IZUMO1 exhibits an elongated rod-shaped structure comprised of a helical bundle IZUMO domain and an immunoglobulin-like domain that are each firmly anchored to an intervening  $\beta$ -hairpin region through conserved disulfide bonds. The central  $\beta$ -hairpin region of IZUMO1 provides the main platform for JUNO binding, while the surface located behind the putative JUNO ligand binding pocket is involved in IZUMO1 binding. Structure-based mutagenesis analysis confirms the biological importance of the

IZUMO1–JUNO interaction. This structure provides a major step towards elucidating an essential phase of fertilization and it will contribute to the development of new therapeutic interventions for fertility, such as contraceptive agents.

Given that only one of many ejaculated spermatozoa (100–300 million in humans) will fertilize an oocyte, several tightly regulated molecular mechanisms must be integrated into the process of fertilization<sup>1–4</sup>. As the culmination of fertilization, gamete membrane fusion in particular necessitates extremely robust regulation. IZUMO1 is a type I transmembrane protein comprised of the IZUMO domain<sup>7</sup>, which contains  $\alpha$ -helices and has been shown to be important for the sperm's oocyte adhesion capability<sup>8</sup>, an immunoglobulin-like domain in the extracellular region, and a short cytoplasmic tail. In contrast, JUNO is a glycosylphosphatidylinositol-anchored folate receptor (FR) family protein that lacks the ability to carry folic acid<sup>6,9</sup> (Extended Data Figs 1 and 2). IZUMO1 and JUNO are ideal targets for contraceptive agents because of their crucial involvement in fertilization.



**Figure 1 | IZUMO1–JUNO interaction.** **a**, Gel-filtration chromatography analysis. Gel-filtration chromatograms (top) and SDS–PAGE analysis stained with Coomassie blue (bottom). For gel source data, see Supplementary Fig. 1. **b**, SV–AUC analysis. The normalized  $c(s)$  distributions were plotted against the sedimentation coefficients  $s_{20,w}$  (S). S, Svedberg (unit of sedimentation coefficient). Estimated molecular

weights are indicated. The observed sedimentation coefficient values for IZUMO1 (2.5 S), JUNO (2.6 S), and the IZUMO1–JUNO complex (3.8 S) corresponded well to the values calculated using the three-dimensional coordinates of the IZUMO1 monomer (2.5 S), JUNO monomer (2.6 S), and 1:1 IZUMO1–JUNO complex (3.7 S), respectively. **c**, Isothermal titration calorimetry analysis. All experiments were conducted at neutral pH.

<sup>1</sup>Graduate School of Pharmaceutical Sciences, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. <sup>2</sup>Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita 565-0871, Japan. <sup>3</sup>Okazaki Institute for Integrative Bioscience, National Institutes of Natural Sciences, 5-1 Higashiyama, Myodaiji, Okazaki, Aichi 444-8787, Japan. <sup>4</sup>Department of Cell Science, Institute of Biomedical Sciences, School of Medicine, Fukushima Medical University, 1 Hikarigaoka, Fukushima City, Fukushima 960-1295, Japan.

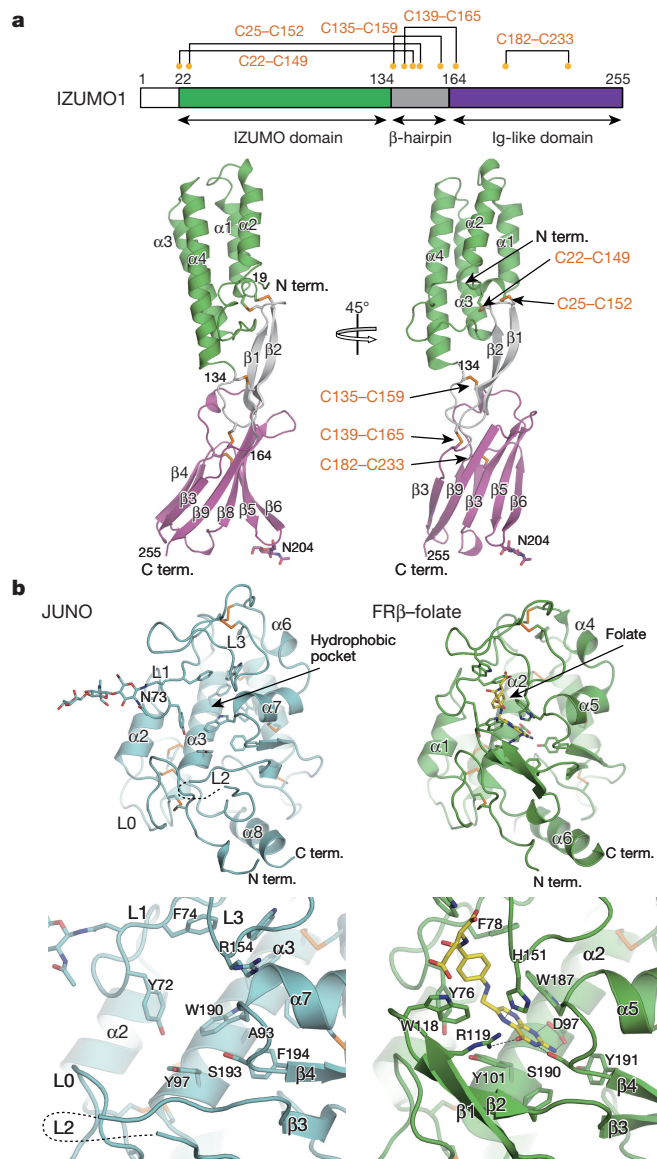
\*These authors contributed equally to this work.

Both the extracellular regions of human IZUMO1 and JUNO alone expressed in S2 cells were monomeric in solution, as confirmed by size-exclusion chromatography and sedimentation velocity analytical ultracentrifugation (SV-AUC) analyses (Fig. 1a, b). Isothermal titration calorimetry analysis showed that IZUMO1 bound JUNO with a  $K_d$  value of 91 nM and a stoichiometry ( $N$ ) of approximately 1.0 at pH 7.5 (Fig. 1c, Extended Data Table 1), suggesting a 1:1 binding mode. The estimated  $M_w$  of the IZUMO1–JUNO complex ( $s_{20,w} = 3.8$  S) corresponded to a 1:1 complex (Fig. 1b). The IZUMO1–JUNO interaction was not affected by deglycosylation (Extended Data Table 1).

We determined the crystal structures of human IZUMO1, JUNO (form 1 and form 2), and the IZUMO1–JUNO complex (form 1, form 2 and form 3) (Extended Data Tables 2 and 3). IZUMO1 exhibits an elongated rod-shaped structure approximately 90 Å in length, comprised of the N-terminal  $\alpha$ -helical IZUMO domain (residues 22–134), a central  $\beta$ -hairpin region (residues 135–163), and the C-terminal immunoglobulin-like domain (residues 164–255) (Fig. 2a, Extended Data Fig. 1). The IZUMO domain forms a four-helix bundle ( $\alpha 1$ – $\alpha 4$ ) possessing up-down-up-down topology, in which pairwise  $\alpha$ -helices ( $\alpha 1$ – $\alpha 2$  and  $\alpha 3$ – $\alpha 4$ ) are arranged in an anti-parallel mode (Fig. 2a). IZUMO1 contains ten cysteines, which are conserved among mammalian species (Extended Data Fig. 1), that are all involved in disulfide bond formation (Fig. 2a). The positions of the disulfide bonds were confirmed in the anomalous difference Fourier maps based on diffraction data collected at a 2.7 Å wavelength (Extended Data Fig. 3). The flanking region of the  $\alpha 1$  helix (C22–XX–C25) is tethered to the loop (C149–XX–C152) of the central  $\beta$ -hairpin region by the disulfide bonds between C22–C149 and C25–C152 (Fig. 2a); the C–XX–C motifs in both regions are conserved among species (Extended Data Fig. 1). Similarly, the C139–C165 disulfide bond stabilizes the interaction between the central  $\beta$ -hairpin region ( $\beta 1$  and  $\beta 2$ ) and the immunoglobulin-like domain (Fig. 2a). The remaining disulfide bonds (C135–C159 and C182–C233) stabilize the conformation of the central region and immunoglobulin-like domain, respectively (Fig. 2a). These disulfide bond-mediated interactions with the central  $\beta$ -hairpin region restrict the relative orientation of the three regions despite the absence of direct interactions between the IZUMO and immunoglobulin-like domains (Extended Data Fig. 4a).

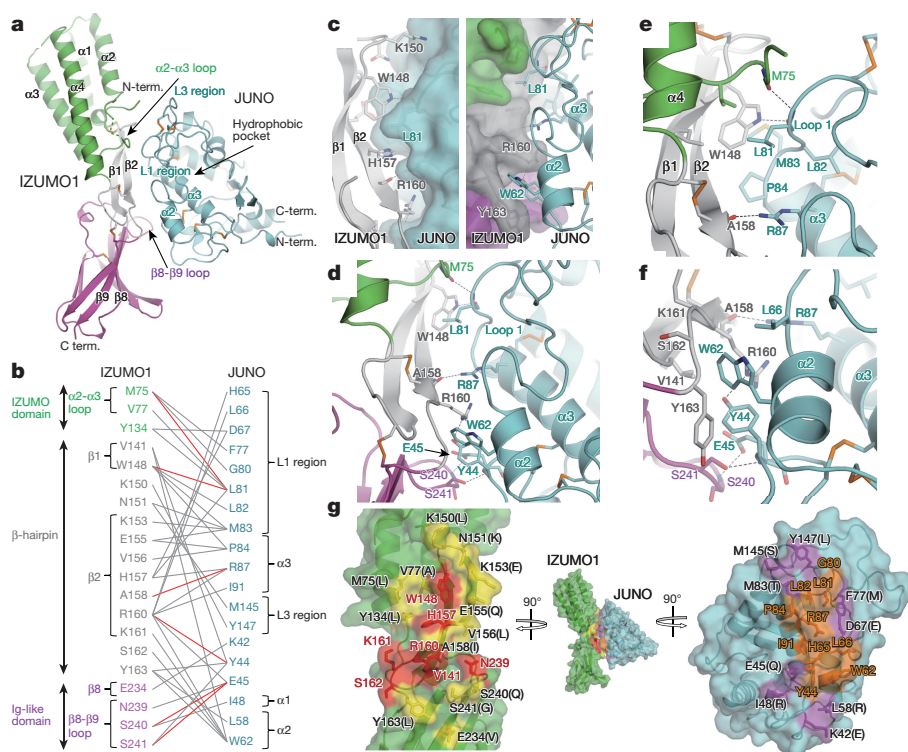
The structure of human JUNO is similar to that of related FRs (Fig. 2b). Eight  $\alpha$ -helices ( $\alpha 1$ – $\alpha 8$ ) and four  $\beta$ -strands ( $\beta 1$ – $\beta 4$ ) in JUNO form a single globular fold, which is stabilized by the eight disulfide bonds conserved among FRs<sup>10,11</sup>, riboflavin-binding protein<sup>12</sup>, and JUNO. Although a number of the loop regions (L0, L1 and L3 regions) are disordered in the mouse JUNO structure<sup>9</sup>, these same loop regions in human JUNO form ordered structures in all molecules in the asymmetric unit (Extended Data Fig. 4b). The human JUNO L2 region, which corresponds to the inhibitory loop in FRs<sup>11</sup>, is disordered (Fig. 2b), possibly due to a 6–7 amino acid insertion not present in other species (Extended Data Figs 1 and 2). Because of the high sequence identity with FRs (~60%), JUNO is also thought to have a hydrophobic pocket that corresponds to the folate binding pocket in FRs (Fig. 2b, Extended Data Fig. 2). The JUNO pocket has a volume of 450 Å<sup>3</sup> and is delimited by six structural segments (L0–L4 regions and  $\alpha 3$  helix). Although most of the hydrophobic residues forming the folate binding pocket are conserved in JUNO (Fig. 2b, Extended Data Fig. 2), it does not exhibit any affinity for folate<sup>6</sup>. In addition to the hydrophobic interactions, FR $\beta$  forms multiple polar interactions with the folate pterin moiety via D97, R119, H151, and S190 (Fig. 2b), which have been shown to be important for folate binding in FRs<sup>10,11</sup>. These residues are not conserved in JUNO except for S190; corresponding residues for JUNO are A93, Q122, R154, and S193, respectively (Fig. 2b, Extended Data Fig. 2). Moreover, the side chain of W190 in JUNO exhibits a rotamer different from that in FRs (W187 for FR $\beta$ ) (Fig. 2b), resulting in narrowing of the pocket. Therefore, JUNO would be unable to bind folate.

IZUMO1–JUNO exists as a 1:1 complex in crystals (Fig. 3a). The three crystal forms of the IZUMO1–JUNO complex revealed essentially



**Figure 2 | Structure of IZUMO1 and JUNO.** **a**, Structure of human IZUMO1. The Cys positions and disulfide bonds are indicated. The disulfide bonds and *N*-acetylglucosamine (NAG) residue are shown as stick representations. An NAG residue attached to N204 is located at the tip of the  $\beta 5$ – $\beta 6$  hairpin region that turns outward from the  $\beta$ -sheet consisting of  $\beta 3$ ,  $\beta 5$ ,  $\beta 6$ ,  $\beta 8$ , and  $\beta 9$ , which form the bottom of IZUMO1. N204 is conserved among species and is important for protecting IZUMO1 from fragmentation in the cauda epididymis<sup>16</sup>. Ig, immunoglobulin. **b**, Structure of human JUNO. Top, structure of human JUNO (left) and the FR $\beta$ –folate complex (PDB ID, 4KMZ)<sup>10</sup> (right). The glycan attached to N73, which is important for JUNO secretion<sup>9</sup>, is located at the limb of the pocket (L1 region). Bottom, magnified views of the JUNO hydrophobic pocket (left) and FR $\beta$  folate binding pocket (right). The hydrogen bonds are depicted with dashed lines.

the same architectures of the complex (Extended Data Fig. 4c), despite the differences in crystallization conditions and protein samples in terms of glycosylation and reductive alkylations of lysine residues (Extended Data Tables 2 and 3). The structures of IZUMO1 and JUNO are not substantially altered upon their forming the complex (Extended Data Fig. 4c). The IZUMO1–JUNO interface is composed of the IZUMO1 regions of the  $\alpha 2$ – $\alpha 3$  loop (IZUMO domain), the central  $\beta$ -hairpin region ( $\beta 1$  and  $\beta 2$ ), and the  $\beta 8$ – $\beta 9$  loop (immunoglobulin-like domain) and the JUNO flanking region of the  $\alpha 1$  helix, the  $\alpha 2$  helix, and the L1 region, as well as the  $\alpha 3$  helix, and the N-terminal side of L3 region (JUNO) (Fig. 3a, b, Extended Data Fig. 1).



**Figure 3 | Structure of the IZUMO1-JUNO complex.** **a**, Structure of the human IZUMO1-JUNO complex. **b**, Schematic summary of IZUMO1-JUNO interactions. Red and grey lines depict hydrogen bonds and van der Waals contacts, respectively. **c**, Surface complementarity of the IZUMO1-JUNO interface. JUNO (left) and IZUMO1 (right) are shown in semi-transparent surface representation. **d**, Hydrogen bond interaction between IZUMO1 and JUNO. **e**, **f**, Trp-mediated interactions

between IZUMO1 and JUNO, focusing on W148 of IZUMO1 (**e**) and W62 of JUNO (**f**). **g**, The interface of IZUMO1 (left) and JUNO (right). The interface residues that are conserved between human and mouse are coloured in red (left) and orange (right), and non-conserved residues in yellow (left) and magenta (right). The residues of mouse IZUMO1 and JUNO are shown in parenthesis.

The central  $\beta$ -hairpin region of IZUMO1 acts as the main platform for the JUNO-binding surface located behind the putative ligand-binding pocket (Fig. 3a, b). IZUMO1 and JUNO utilize their complementary surface for association, resulting in a contact area of 842 Å<sup>2</sup> (crystal form 1) (Fig. 3c). Hydrophobic and van der Waals interactions are the main contributors to the binding; in addition, six hydrogen bonds also contribute to the binding (Fig. 3d). Interactions through the two Trp residues present on each protein with good surface complementarity can be defined: the side chain of IZUMO1 W148 interacts with L81, L82, M83, and P84 of JUNO (Fig. 3c left, e); and JUNO W62 interacts with R160, K161, S162, and Y163 of IZUMO1 (Fig. 3e, right, f). These Trp residues are conserved among all species (Extended Data Fig. 1). The L1 and L3 regions of JUNO have been shown to be important for interaction with IZUMO1<sup>9</sup>. The conserved G80-L81 motif in the L1 region undergoes a conformational change upon IZUMO1 binding (Extended Data Figs 1 and 4c), thus creating extensive contacts with M75, V77, Y134, W148, and H157 of IZUMO1 (Fig. 3c, right, e). Accordingly, mutant proteins of the interface residues, especially W148A of IZUMO1 and W62A and L81A of JUNO, exhibited reduced affinity (Fig. 1c, Extended Data Table 1).

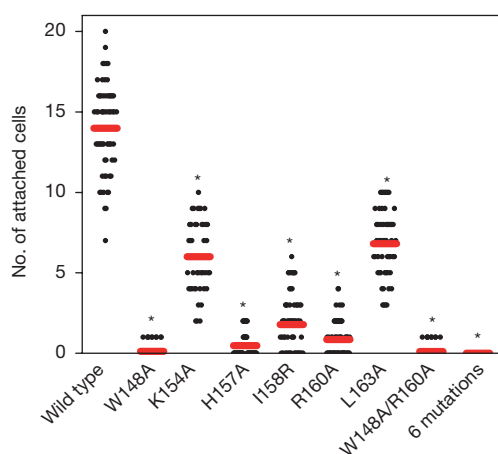
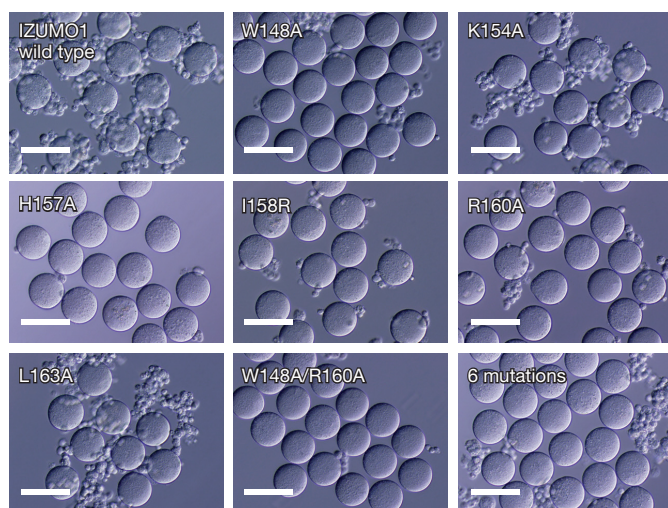
Together with our paper, another group has also reported essentially the same results<sup>13</sup>. Their work further confirms the overall architecture of the IZUMO1-JUNO complex and key residues of the interaction by X-ray crystallography along with techniques complementary to ours, including small angle X-ray scattering, deuterium exchange mass spectrometry, biolayer interferometry, and surface plasmon resonance analyses.

To confirm the functional relevance of the IZUMO1-JUNO interface in sperm-oocyte interactions, we conducted cell-oocyte binding assays using COS-7 cells expressing wild-type or mutant mouse *Izumo1* containing single or multiple mutations in the binding

interface for JUNO (W148A, K154A, H157A, I158R, R160A, and L163A) (Fig. 3b). The cell surface distribution of the proteins was confirmed by immunostaining (Extended Data Fig. 5). The average number of attached cells was significantly reduced by introduction of single mutation in *Izumo1* (Fig. 4). In particular, mutations in conserved amino acid residues (W148A, H157A, and R160A) exhibited the greatest impact on the oocyte binding consistent with the  $K_d$  values of the IZUMO1 W148A mutant (Fig. 1c) and structural observations (Fig. 3). In multiple mutations, COS-7 cells could no longer bind to egg surface, thus demonstrating the functional importance of the IZUMO1-JUNO interface.

The IZUMO1-JUNO interaction has a certain degree of species specificity. For example, human IZUMO1 can interact with hamster JUNO but not with mouse JUNO<sup>6,9,14</sup>. The IZUMO1-JUNO interface includes both conserved and non-conserved residues (Fig. 3g, Extended Data Fig. 1). While the conserved structural features such as the Trp-mediated interactions described above ensure the conserved binding mode, the variable regions in the interface determine the species specificity of the IZUMO1-JUNO interaction.

It has been hypothesized that after the initial IZUMO1-JUNO binding, IZUMO1 forms a closed dimer and simultaneously releases JUNO<sup>15</sup>. Although IZUMO1 tends to dimerize at high concentrations, no further oligomerization of the IZUMO1-JUNO complex in solution was detected by SV-AUC analysis, even at high concentrations (~100  $\mu$ M) (Extended Data Fig. 6). Thus, the current IZUMO1-JUNO structure represents the initial gamete recognition state and further structural conversion of the complex would occur during fertilization. The conformation of the central region of IZUMO1 important for JUNO binding is physically restricted by neighbouring domains through partially exposed disulfide bonds (Fig. 2a). Accordingly, IZUMO1 exhibited high sensitivity to reducing agents and easily lost



**Figure 4 | IZUMO1 containing interface mutations abolishes cell–oocyte binding.** Cell–oocyte assays to evaluate mouse IZUMO1–JUNO interactions were performed using transfected COS-7 cells. Bright-field images taken 1 h post incubation. Scale bars, 100  $\mu$ m. Attached cell count values (excluding aggregated cells) from three independent experiments are presented (bottom). The red line indicates the average. \* $P < 0.0001$  between wild-type IZUMO1 and its mutants; Student's  $t$ -test. Sample size (no. of oocytes), average and error bars are discussed further in the Methods.

its native conformation, while JUNO did not (Extended Data Fig. 7a). Interestingly, the IZUMO1–JUNO interaction was weaker at acidic pH than at neutral pH (Extended Data Fig. 7b, Extended Data Table 1). As FRs undergo pH-dependent conformational change<sup>10</sup>, JUNO may also exhibit a distinct conformation at low pH that is incapable of binding to IZUMO1. These properties of IZUMO1 and JUNO may be involved in the regulation of the IZUMO1–JUNO interaction. Alternatively, unidentified factor(s) on the oocyte could be involved in the subsequent fertilization steps. Additional studies will be required to elucidate the processes following the initial encounter of IZUMO1 and JUNO.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 February; accepted 20 May 2016.

Published online 15 June 2016.

1. Klinovska, K., Sebkova, N. & Dvorakova-Hortova, K. Sperm–egg fusion: a molecular enigma of mammalian reproduction. *Int. J. Mol. Sci.* **15**, 10652–10668 (2014).
2. Okabe, M. The cell biology of mammalian fertilization. *Development* **140**, 4471–4479 (2013).
3. Evans, J. P. Sperm–egg interaction. *Annu. Rev. Physiol.* **74**, 477–502 (2012).
4. Knobil, E. & Neill, J. D. *The Physiology of Reproduction*. 2nd edn, (Raven Press, 1994).
5. Inoue, N., Ikawa, M., Isotani, A. & Okabe, M. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* **434**, 234–238 (2005).
6. Bianchi, E., Doe, B., Goulding, D. & Wright, G. J. Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **508**, 483–487 (2014).
7. Ellerman, D. A. *et al.* Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. *Mol. Reprod. Dev.* **76**, 1188–1199 (2009).
8. Inoue, N. *et al.* Molecular dissection of IZUMO1, a sperm protein essential for sperm–egg fusion. *Development* **140**, 3221–3229 (2013).
9. Han, L. *et al.* Divergent evolution of vitamin B9 binding underlies Juno-mediated adhesion of mammalian gametes. *Curr. Biol.* **26**, R100–R101 (2016).
10. Wibowo, A. S. *et al.* Structures of human folate receptors reveal biological trafficking states and diversity in folate and antifolate recognition. *Proc. Natl Acad. Sci. USA* **110**, 15180–15188 (2013).
11. Chen, C. *et al.* Structural basis for molecular recognition of folic acid by folate receptors. *Nature* **500**, 486–489 (2013).
12. Monaco, H. L. Crystal structure of chicken riboflavin-binding protein. *EMBO J.* **16**, 1475–1483 (1997).
13. Aydin, H., Sultana, A., Li, S., Thavalingam, A. & Lee, J. E. molecular architecture of the human sperm Izumo1 and egg Juno fertilization complex. *Nature* <http://dx.doi.org/10.1038/nature18595> (this issue).
14. Bianchi, E. & Wright, G. J. Cross-species fertilization: the hamster egg receptor, Juno, binds the human sperm ligand, Izumo1. *Phil. Trans. R. Soc. B* **370** (2015).
15. Inoue, N., Hagihara, Y., Wright, D., Suzuki, T. & Wada, I. Oocyte-triggered dimerization of sperm IZUMO1 promotes sperm–egg fusion in mice. *Nat. Commun.* **6**, 8858 (2015).
16. Inoue, N., Ikawa, M. & Okabe, M. Putative sperm fusion protein IZUMO and the role of N-glycosylation. *Biochem. Biophys. Res. Commun.* **377**, 910–914 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the beamline staff members at the Photon Factory and SPring-8 for their assistance with data collection. We thank Y. Yamada and D. Liebschner for assistance with S-SAD data collection at PF-1A. This research is supported by the Platform Project for Supporting in Drug Discovery and Life Science Research (Platform for Drug Discovery, Informatics, and Structural Life Science) from the Japan Agency for Medical Research and Development (AMED). This work was supported by a Grant-in-Aid from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (U.O., S.U., N.I. and T.S.); CREST, JST (T.S.); the Takeda Science Foundation (U.O., N.I. and T.S.); the Mochida Memorial Foundation for Medical and Pharmaceutical Research (U.O.); and the Daiichi Sankyo Foundation of Life Science (U.O.).

**Author Contributions** H.I. expressed and purified recombinant proteins and performed size-exclusion chromatography and isothermal titration calorimetry experiments. H.I. and U.O. performed crystallization and structure determination. E.K. and S.U. performed AUC analyses. N.I. performed cellular assays. U.O. and T.S. directed the research and wrote the paper with assistance from all other authors.

**Author Information** The coordinates and structure-factor data of human IZUMO1, JUNO (form 1 and form 2), IZUMO1–JUNO complex (form 1, form 2, and form 3) have been deposited in the Protein Data Bank under the accession numbers 5JK9, 5JKA, 5JKB, 5JKC, 5JKD and 5JKE, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to U.O. ([umeji@mol.f.u-tokyo.ac.jp](mailto:umeji@mol.f.u-tokyo.ac.jp)) or T.S. ([shimizu@mol.f.u-tokyo.ac.jp](mailto:shimizu@mol.f.u-tokyo.ac.jp)).

**Reviewer Information** *Nature* thanks K. Melcher, M. Okabe and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

**Protein expression, purification and crystallization.** The DNA encoding human IZUMO1 (residues 22–255, Uniprot accession number Q8IYV9) and human JUNO (residues 20–228, Uniprot accession number A6ND01) with a C-terminal PreScission protease cleavage site followed by Protein A tag were inserted into the expression vector pMT/BiP/V5-His of *Drosophila* Expression System. *Drosophila* S2 cells were co-transfected with the IZUMO1 or JUNO, and pCoHygro vectors. Stably transfected cells were selected in Sf-900 II SFM medium containing  $300 \mu\text{g ml}^{-1}$  hygromycin. Protein expression was induced by adding  $0.5 \text{ mM}$   $\text{CuSO}_4$  in Express Five SFM medium. Culture supernatant was harvested at 160–240 h after induction. Protein was captured by IgG Sepharose 6 Fast Flow (GE Healthcare) equilibrated with phosphate buffered saline (PBS), washed with ten column volumes of PBS, and eluted by  $0.1 \text{ M}$  glycine-HCl pH 3.5 and  $0.15 \text{ M}$  NaCl. Eluent was immediately neutralized by adding with 1/20 volume of  $1 \text{ M}$  Tris-HCl pH 8.0 and was concentrated to  $1\text{--}3 \text{ mg ml}^{-1}$ . For JUNO, protein was purified by HiTrap Q (GE Healthcare) anion exchange chromatography at pH 8.5. Bound protein was eluted with a linear NaCl gradient ( $0$  to  $0.6 \text{ M}$ ). Protein was incubated with 1/20 (w/w) PreScission protease for 3 h at  $277 \text{ K}$  to cleave Protein A tag and further purified by Superdex 200 gel filtration column (GE healthcare) equilibrated with  $10 \text{ mM}$  Tris-HCl pH 7.5 and  $0.15 \text{ M}$  NaCl. For the preparations of deglycosylated samples of IZUMO1 and JUNO, the culture medium was supplemented with kifunensine ( $1.8 \mu\text{g ml}^{-1}$ ) to produce protein with endoglycosidase-susceptible *N*-glycans and the purification steps included deglycosylation. After tag cleavage, protein was added with 1/10 volume of  $1 \text{ M}$  MES-NaOH pH 6.5 and  $1\text{--}2 \text{ U}$  Endo H<sub>f</sub> (New England Biolabs) per mg of protein and incubated for 3 h at room temperature. The crystallization samples with reductive alkylation of lysine residues were prepared using Reductive Alkylation Kit (Hampton Research) according to the manufacturer's protocol.

Crystallization experiments were performed with sitting-drop vapour-diffusion methods at  $293 \text{ K}$ . The crystallization droplets were made by mixing the equivolume of protein solution and reservoir solution, typically around  $0.2\text{--}1.0 \mu\text{l}$ . The crystallization conditions are summarized in Extended Data Table 2. The Os derivative of IZUMO1 crystals were prepared by soaking the IZUMO1 crystals into a solution containing  $10 \text{ mM}$   $\text{K}_2\text{OsO}_4$  in mother liquor ( $12\%$  (w/v) PEG6000,  $0.1 \text{ M}$  sodium citrate pH 5.6,  $90 \text{ mM}$  NaCl) for 1 h. The Pt derivative of IZUMO1 crystals were prepared by soaking the IZUMO1 crystals into a solution containing  $10 \text{ mM}$   $\text{K}_2\text{PtCl}_4$  in the mother liquor for 1 minute.

**Data collection and structure determination.** Diffraction data sets were collected on beamlines PF-1A and PF-AR NE3A (Ibaraki, Japan) under cryogenic conditions at  $100 \text{ K}$ . Crystals were soaked into cryoprotectant solution (Extended Data Table 2) and then flash-cooled under a cold gas stream. The diffraction data sets were processed using the HKL2000 package<sup>17</sup> or XDS<sup>18</sup>. Phasing and initial model building of the IZUMO1 crystal structure were performed using autoSHARP<sup>19</sup> and ARP/wARP<sup>20</sup>, respectively, followed by iterative cycles of manual model building using COOT program<sup>21</sup> and restrained refinement using REFMAC<sup>22</sup> until the *R* factor was converged. The initial models for JUNO structure (form 1) was obtained by the molecular replacement method using the MOLREP<sup>23</sup> program using the coordinates of FR $\beta$  (PDB ID: 4KMY) and was further refined similarly to the IZUMO1 structure. The initial models for the remaining structures were obtained by the molecular replacement method using the refined coordinates of IZUMO1 and JUNO (form 1). The quality of the final structures was evaluated with MolProbity<sup>24</sup>. The statistics of the data collection and refinement are summarized in Extended Data Table 3. Figures were prepared with PyMOL<sup>25</sup>. The pocket volume of JUNO was calculated using CASTp<sup>26</sup>.

**Isothermal titration calorimetry.** ITC experiments were performed at  $298 \text{ K}$  using MicroCal iTC200 (GE Healthcare) in a buffer composed of  $10 \text{ mM}$  Tris-HCl pH 7.5 and  $0.15 \text{ M}$  NaCl (Fig. 1c) or  $10 \text{ mM}$  MES-NaOH pH 5.5 and  $0.15 \text{ M}$  NaCl (Extended Data Fig. 7b). JUNO at a concentration of  $100 \mu\text{M}$  was titrated into  $10 \mu\text{M}$  of IZUMO1. The titration sequence included a single  $0.4 \mu\text{l}$  injection followed by 19 injections of  $2 \mu\text{l}$  each, with a spacing of 120 s between injections. OriginLab software (GE Healthcare) was used to analyse the raw ITC data. Thermodynamic parameters were extracted from curve fitting analysis with a single-site binding model.

**Sedimentation velocity analytical ultracentrifugation (SV-AUC).** SV-AUC experiments were performed at  $20^\circ\text{C}$  in a ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter) at  $42,000 \text{ r.p.m.}$  using absorbance detection. The collected data were analysed using continuous *c*(s) distribution of SEDFIT<sup>27,28</sup> fitting for the frictional ratio, meniscus, and time-invariant noise and using

regularization level of 0.68. The buffer density and viscosity and the partial specific volumes of IZUMO1 and JUNO were calculated using the program SEDNTERP 1.09 and were  $1.00852 \text{ g ml}^{-1}$  and  $1.0256 \text{ cP}$ ,  $0.7217 \text{ cm}^3 \text{ g}^{-1}$ , and  $0.7349 \text{ cm}^3 \text{ g}^{-1}$ , respectively. The partial specific volume of IZUMO1–JUNO complex was estimated using the program UltraScan-SOMO<sup>29</sup> and was  $0.728 \text{ cm}^3 \text{ g}^{-1}$ .

To study the complex formation between IZUMO1 and JUNO, experiments were conducted with  $20 \mu\text{M}$  of each protein individually and as a mixture (Fig. 1b). The concentration-dependent dimerization of IZUMO1 was revealed through measurements performed at protein concentrations of  $10$ ,  $20$  and  $100 \mu\text{M}$  (Extended Data Fig. 6a). The absence of concentration dependence of IZUMO1–JUNO complex formation was demonstrated using mixtures containing  $10$ ,  $20$  and  $100 \mu\text{M}$  of each respective protein (Extended Data Fig. 6b). In addition, SV-AUC experiments were conducted for the mixture of deglycosylated proteins (Extended Data Fig. 6b). Theoretical values of sedimentation coefficients of IZUMO1, JUNO and the IZUMO1–JUNO complex were calculated from the three-dimensional structures using the program UltraScan-SOMO<sup>29</sup>.

**Mice, antibodies and cultured cells.** Eight- to twelve-week-old B6D2F1 (a cross between female C57BL/6 and male DBA/2) female mice were purchased from Japan SLC, Inc. All animal experiments were approved by the Animal Care and Use Committee of Fukushima Medical University, Japan. Anti-mouse IZUMO1 monoclonal antibody (Mab18), generated as described<sup>15</sup>, was conjugated with Alexa Fluor 488 using an antibody labelling kit (Thermo Fisher Scientific). COS-7 cells (African green monkey kidney fibroblast-like cell line) were obtained from RIKEN BRC. This cell line has been authenticated mycoplasma free by PCR.

**Cell-oocyte assay.** Mouse *Izumo1* and its mutants (W148A, K154A, H157A, I158R, R160A, L163A, W148A/R160A, and the 6 mutant (W148A/K154A/H157A/I158R/R160A/L163A)), which were created using KOD-plus-neo Mutagenesis (Toyobo), cDNAs were ligated into the mammalian expression vector pCXN-2. These constructs were verified by DNA sequencing. COS-7 cells were transiently transfected with these constructs using polyethylenimine (PEI) methods<sup>30</sup>. To verify cell surface expression of IZUMO1 proteins (Extended Data Fig. 5), IZUMO1 was detected with a mouse IZUMO1-specific monoclonal antibody, Mab18 conjugated with Alexa Fluor 488 (green) without plasma membrane permeabilization treatments. The final concentration of antibodies added was  $0.5 \mu\text{g ml}^{-1}$ . Nuclei were stained with Hoechst 33342. After two days, transfected COS-7 cells were collected with  $5 \text{ mM}$  EDTA-PBS, washed three times with PBS, and suspended in TYH medium (LSI Medience). For preparation of zona-free eggs, B6D2F1 female mice ( $>8$  weeks old) were superovulated with an injection of  $7.5 \text{ IU}$  of human chorionic gonadotropin (hCG)  $48 \text{ h}$  after a  $7.5\text{-IU}$  injection of equine chorionic gonadotropin (eCG). The eggs were collected from the oviduct  $16 \text{ h}$  after the hCG injection. Eggs were placed in a  $200\text{-}\mu\text{l}$  drop of TYH medium. The zona pellucida was removed from eggs by treatment with  $1.0 \text{ mg ml}^{-1}$  of collagenase (Wako). Zona-free eggs were incubated with transfected COS-7 cells at  $37^\circ\text{C}$  in TYH medium for  $1 \text{ h}$  and the attached cells were counted under an inverted microscope after briefly washing by pipetting (68 oocytes for wild-type mouse IZUMO1 (average attached cells:  $13.99 \pm 0.32$ ), 82 oocytes for W148A (average attached cells:  $0.12 \pm 0.04$ ), 54 oocytes for K154A (average attached cells:  $6.0 \pm 0.27$ ), 72 oocytes for H157A (average attached cells:  $0.38 \pm 0.08$ ), 69 oocytes for I158R (average attached cells:  $1.78 \pm 0.21$ ), 181 oocytes for R160A (average attached cells:  $0.82 \pm 0.13$ ), 54 oocytes for L163A (average attached cells:  $6.81 \pm 0.28$ ), 84 oocytes for W148A/R160A (average attached cells:  $0.1 \pm 0.04$ ), and 71 oocytes for W148A/K154A/H157A/I158R/R160A/L163A (average attached cells: 0); three independent experiments).

17. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

18. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).

19. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).

20. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protocols* **3**, 1171–1179 (2008).

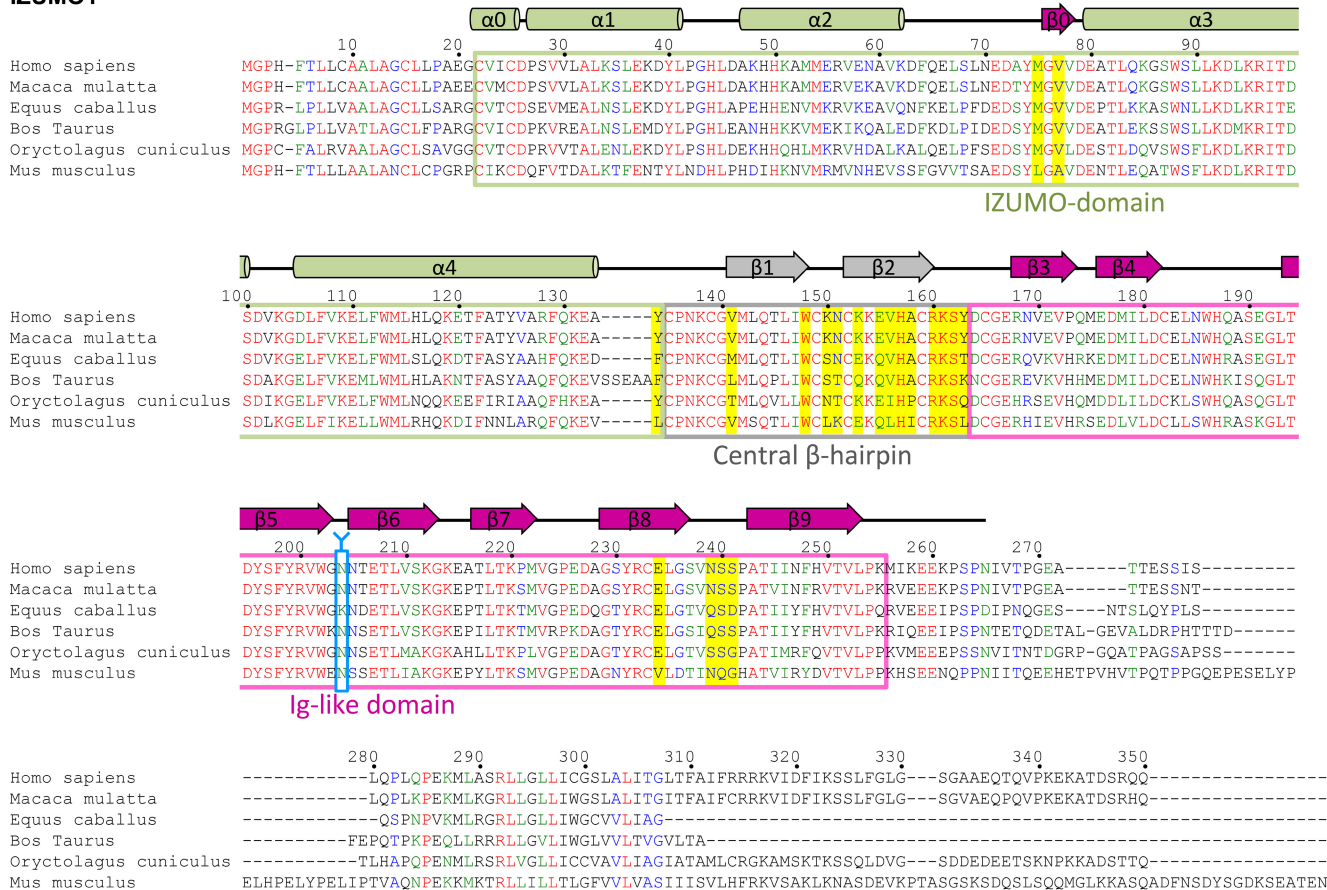
21. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

22. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).

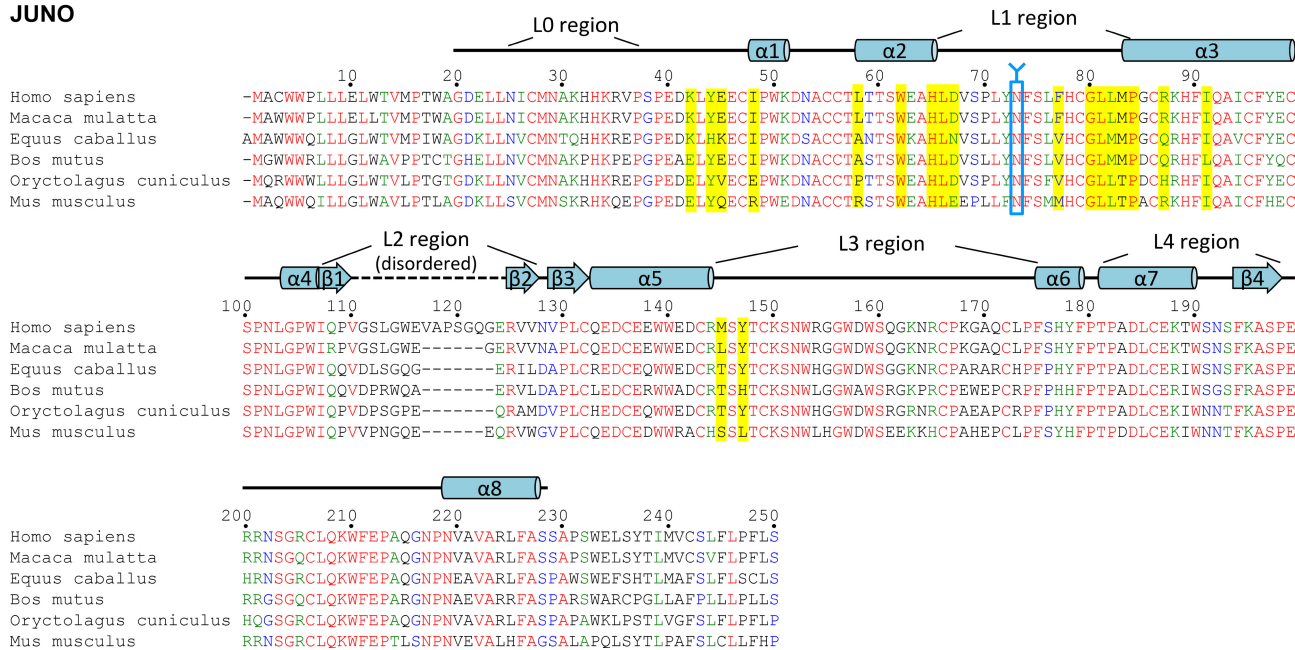
23. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).

24. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
25. DeLano, W. L. The PyMOL Molecular Graphics System. <http://www.pymol.org> (2008).
26. Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–W118 (2006).
27. Brautigam, C. A. Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol.* **562**, 109–133 (2015).
28. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
29. Brookes, E., Demeler, B., Rosano, C. & Rocco, M. The implementation of SOMO (SOLution MODeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *Eur. Biophys. J.* **39**, 423–435 (2010).
30. Hashiguchi, T. *et al.* Crystal structure of measles virus hemagglutinin provides insight into effective vaccines. *Proc. Natl Acad. Sci. USA* **104**, 19535–19540 (2007).

## IZUMO1

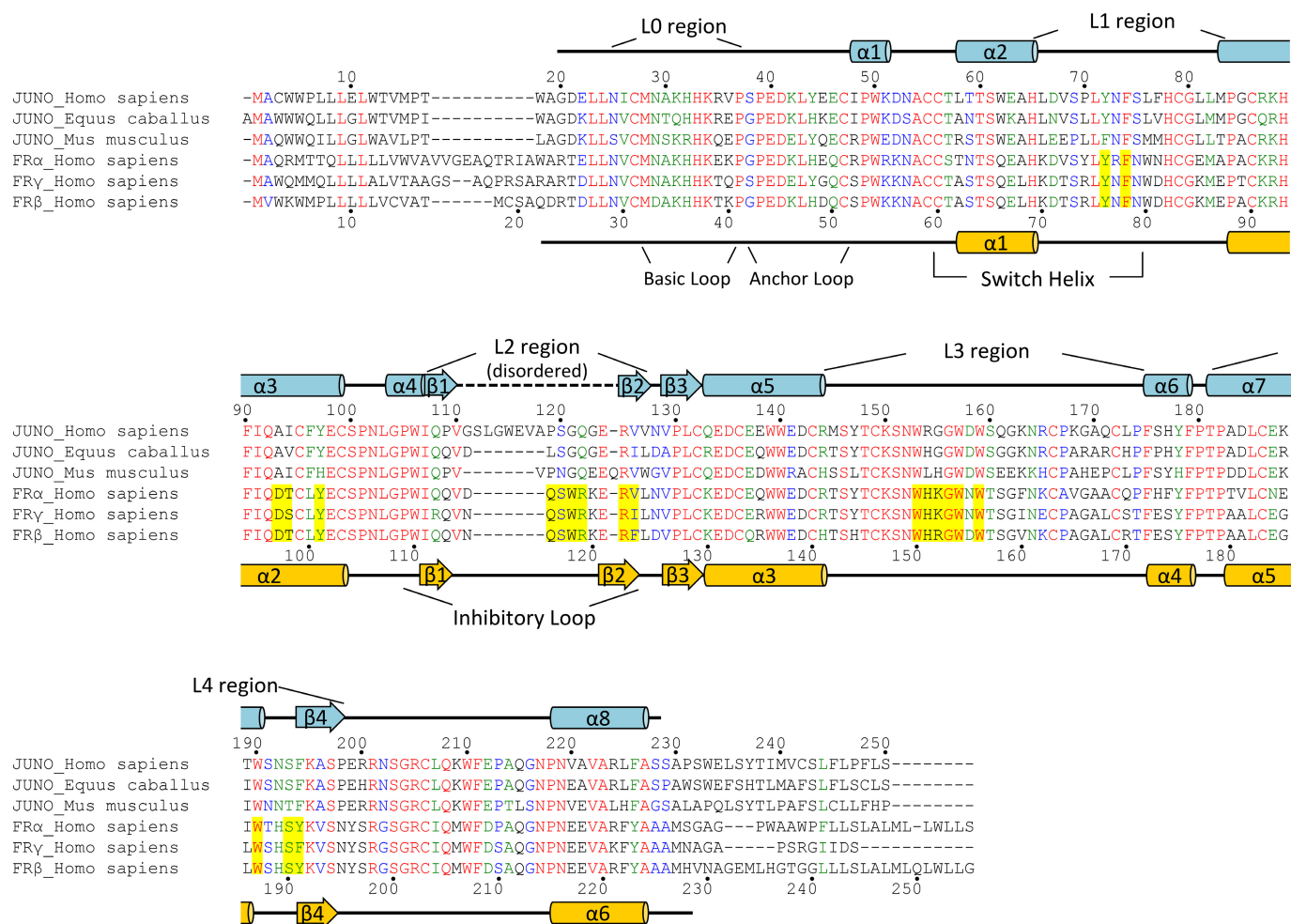


## JUNO



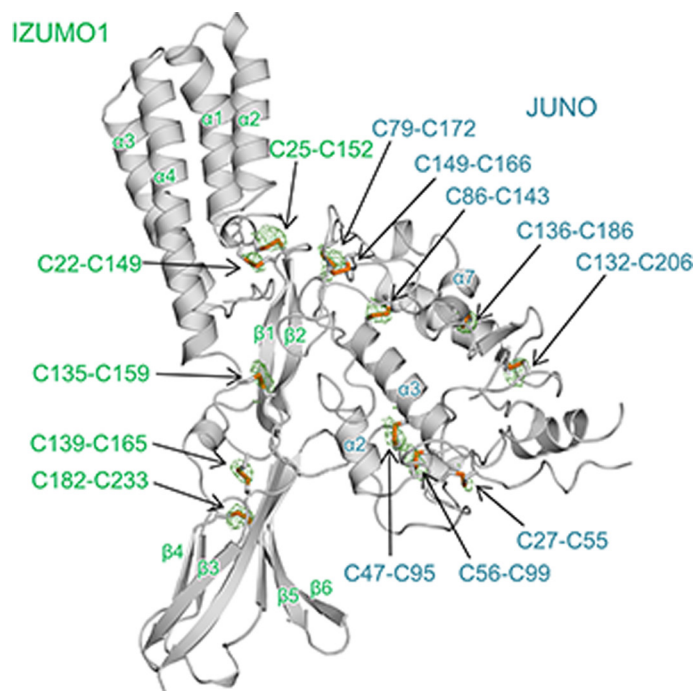
**Extended Data Figure 1 | Sequence alignments of IZUMO1 and JUNO.** Sequence alignments of IZUMO1 (top) and JUNO (bottom) from *Homo sapiens*, *Macaca mulatta*, *Equus caballus*, *Bos taurus* or *Bos mutus*, *Oryctolagus cuniculus* and *Mus musculus* are shown. Secondary structure elements are displayed above the sequences. The residues of the IZUMO1–JUNO interface are indicated by yellow highlighting. The sequence identity between human and mouse IZUMO1 and JUNO are 59% and 69%, respectively, for the extracellular regions used in this study.

In contrast, the interface residues exhibit less conservation (35% and 53% identity for IZUMO1 and JUNO, respectively). The N-glycosylation sites are indicated by blue Y-shaped characters. Alignments were performed using Clustal Omega software (EMBL–European Bioinformatics Institute). Residues are coloured to indicate the degree of similarity: red residues are those with the highest similarity, followed by green, blue, and black (lowest similarity).

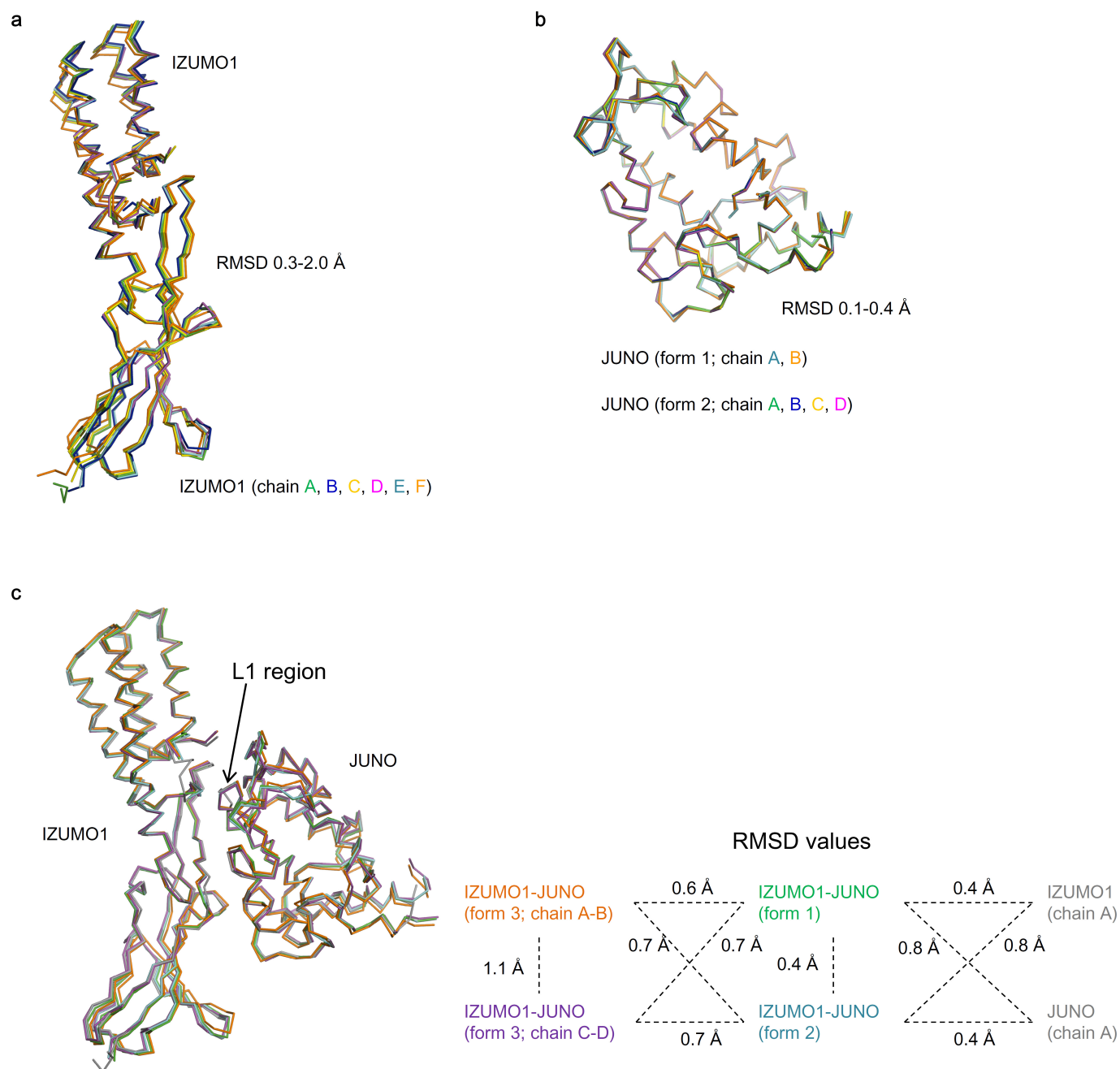


**Extended Data Figure 2 | Sequence alignment of JUNO and folate receptor (FR).** Sequence alignment of JUNO from *Homo sapiens*, *Equus caballus* and *Mus musculus* and FR $\alpha$ , FR $\beta$ , and FR $\gamma$  from *Homo sapiens*. Secondary structure elements for JUNO and FR are displayed above and below the sequences, respectively. The residues involved in the

folate binding in FRs are indicated by yellow highlighting. Alignments were performed using Clustal Omega software (EMBL – European Bioinformatics Institute). Residues are coloured to indicate the degree of similarity: red residues are those with the highest similarity, followed by green, blue, and black (lowest similarity).

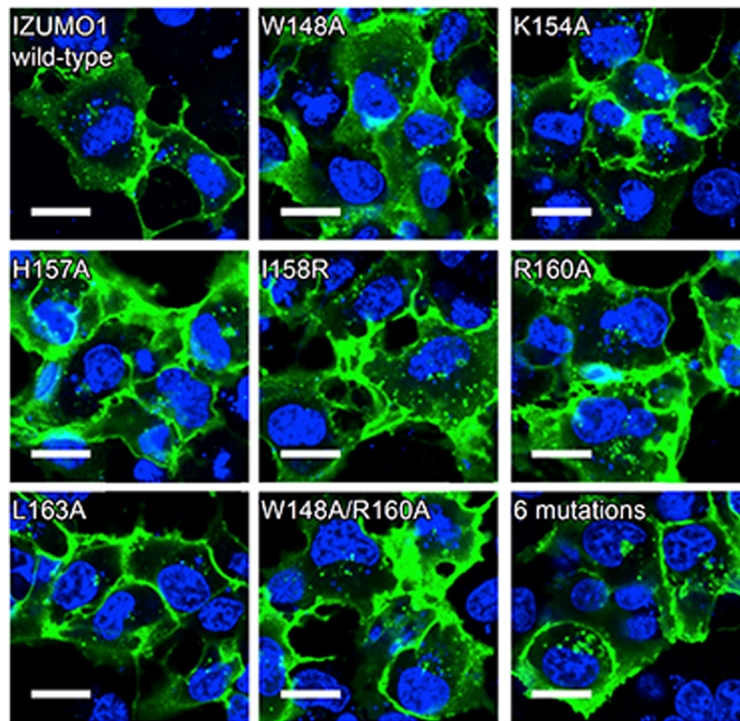


**Extended Data Figure 3 | Anomalous difference Fourier maps from S-SAD data.** The anomalous difference Fourier maps from the data collected at 2.7 Å wavelength contoured with green at the  $3\sigma$  level are superposed onto the refined model of the IZUMO1–JUNO complex. The disulfide bonds are shown in stick representations and indicated by green (IZUMO1) and cyan (JUNO) labels.

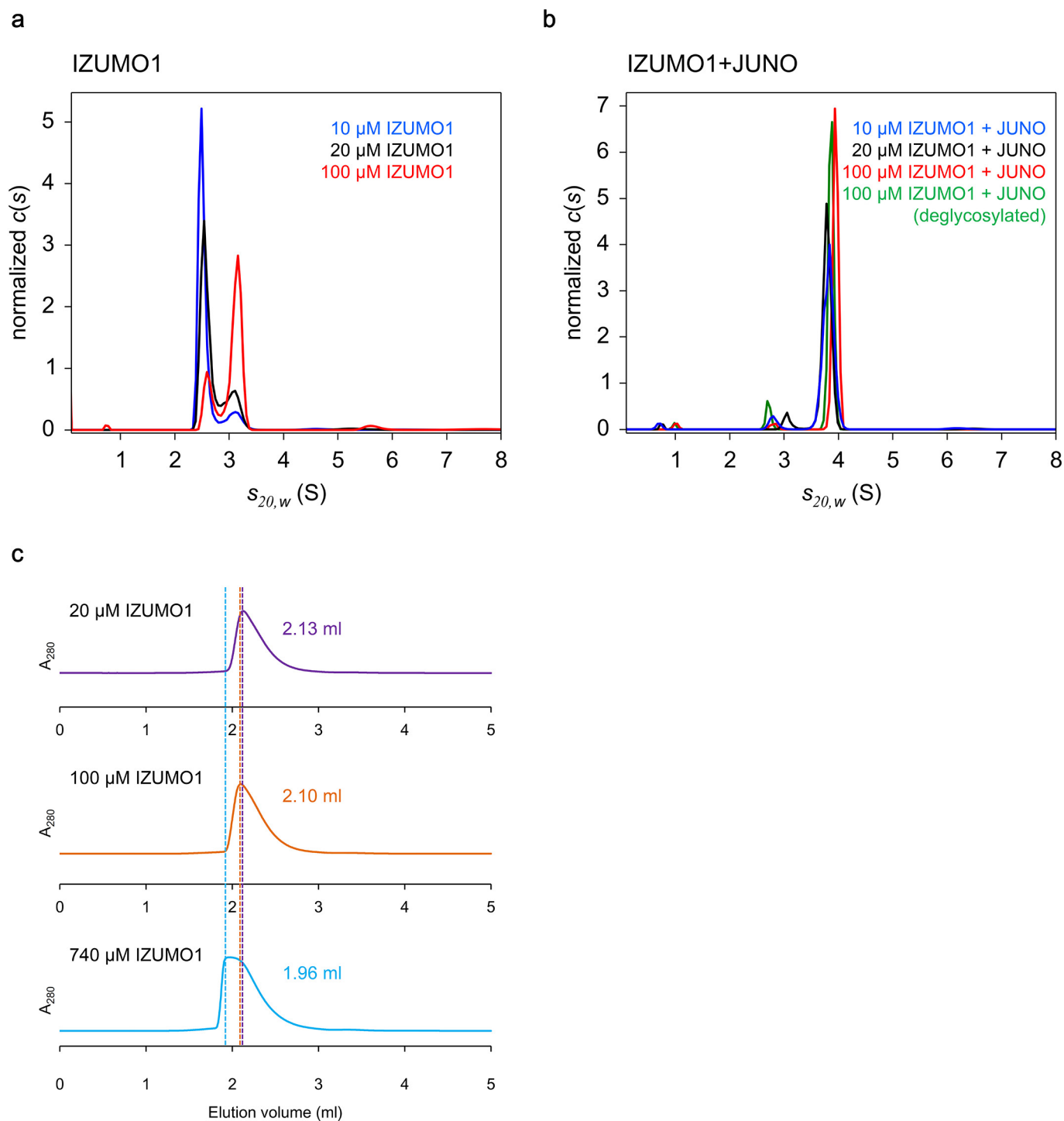


**Extended Data Figure 4 | Superpositions of IZUMO1 and JUNO structures.** **a**, Superposition of the six IZUMO1 molecules (chains A-F) in the asymmetric unit. **b**, Superposition of the two (form 1; chains A and B) and four (form 2; chains A-D) JUNO molecules in the asymmetric unit. **c**, Superposition of IZUMO1 (chain A) and JUNO (chain A) structures

onto the IZUMO1-JUNO complex structure (form 1, form 2, and form 3). IZUMO1 (chain A) and JUNO (chain A) are coloured grey. The IZUMO1-JUNO complexes in the form 1, form 2, and form 3 (chain A-B and chain C-D) crystals are coloured green, cyan, orange, and magenta, respectively. The r.m.s.d. values for each pair of molecules are shown schematically.

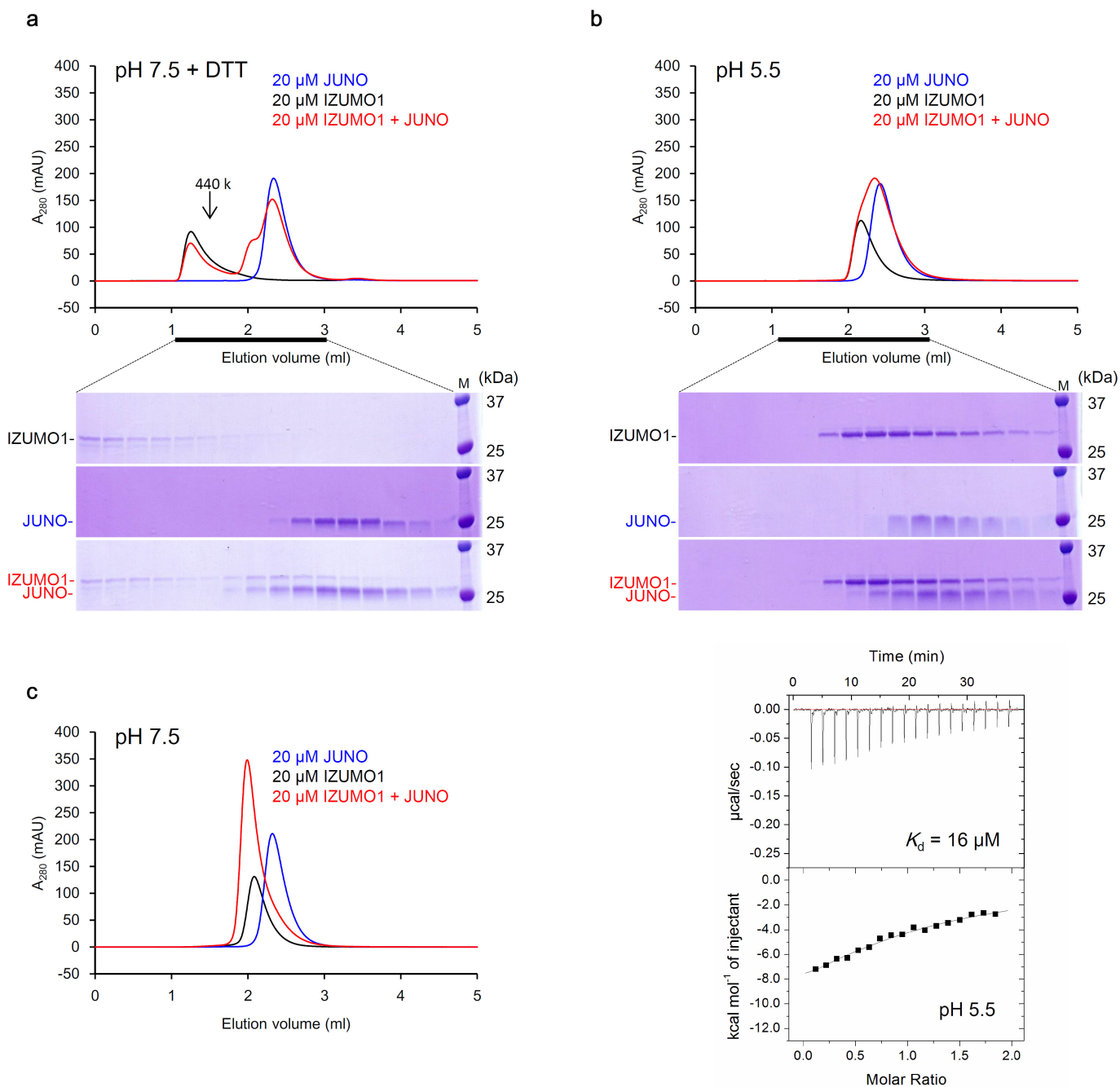


**Extended Data Figure 5 | Cell surface expression of IZUMO1 mutants in COS-7 cells (related to Fig. 4).** Surface localization of wild-type IZUMO1 or its mutant in COS-7 cells. IZUMO1 proteins on the cell surface (green) and nuclei (blue) were stained. Scale bars, 20 μm.



**Extended Data Figure 6 | Dimerization of IZUMO1 and IZUMO1–JUNO complex analysed by SV–AUC and size-exclusion chromatography.** **a, b,** The oligomerization states of IZUMO1 (**a**) and IZUMO1–JUNO complex (**b**) were analysed by SV–AUC at various concentrations. The normalized  $c(s)$  distributions were plotted against the sedimentation coefficients,  $s_{20,w}$  (S). The observed sedimentation coefficient of the IZUMO1 dimer (3.2 S) at a high concentration

(100  $\mu\text{M}$ ) in **a** was smaller than the expected value for IZUMO1 dimer (3.7 S), owing to the fast monomer–dimer interconversion kinetics. **c,** The oligomerization of IZUMO1 was analysed by gel-filtration chromatography. In each experiment, 20, 100 or 740  $\mu\text{M}$  IZUMO1 (total volume of 50  $\mu\text{l}$ ) was injected into a Superdex 200 Increase 5/150 GL gel-filtration column (running buffer; 10 mM Tris–HCl pH 7.5 and 150 mM NaCl).



**Extended Data Figure 7 | Effect of reducing agent and acidic pH on the IZUMO1-JUNO interaction.** **a**, DTT-induced aggregation of IZUMO1. Each sample was incubated at room temperature for 2 h in the presence of 1 mM DTT, and then injected into a Superdex 200 Increase 5/150 GL gel-filtration column (running buffer; 10 mM Tris-HCl pH 7.5 and 150 mM NaCl, 1 mM DTT). Eluents were analysed by SDS-PAGE. For gel source data, see Supplementary Fig. 1. **b**, The interaction between IZUMO1 and

JUNO at acidic pH. The IZUMO1-JUNO interaction at acidic pH was analysed by size-exclusion chromatography (SEC; top) and isothermal titration calorimetry (bottom) as in Fig. 1. In SEC analysis, each sample was injected into a Superdex 200 Increase 5/150 GL gel-filtration column (running buffer; 10 mM MES-NaOH pH 5.5 and 150 mM NaCl). Eluents were analysed by SDS-PAGE. **c**, The SEC chromatograms at a neutral pH of 7.5 is shown as a control experiment.

**Extended Data Table 1 | Isothermal titration calorimetry results**

Cell	pH	Titration	$K_d$ (nM)	$\Delta H$ (kcal/mol)	$\Delta S$ (cal/mol/K)	$N$
IZUMO1 (WT)	7.5	JUNO (WT)	$91 \pm 11$	$-12 \pm 0.2$	$-7.9 \pm 0.8$	$1.0 \pm 0.0$
IZUMO1 (WT)	7.5	JUNO (W62A)	360	-7.8	3.2	1.1
IZUMO1 (WT)	7.5	JUNO (L66A)	130	-9.2	0.7	1.2
IZUMO1 (WT)	7.5	JUNO (L81A)	3,300	-4.1	11.3	$1.0^\dagger$
IZUMO1 (WT)	7.5	JUNO (M83A)	240	-10.9	-6.3	1.0
IZUMO1 (W148A)	7.5	JUNO (WT)	3,200	-0.8	22.6	$1.0^\dagger$
IZUMO1 (V156A)	7.5	JUNO (WT)	89	-14.6	-16.7	0.6
IZUMO1 (WT) (deglycosylated)	7.5	JUNO (WT) (deglycosylated)	$110 \pm 45$	$-9.2 \pm 0.5$	$1.1 \pm 0.9$	$1.1 \pm 0.0$
IZUMO1 (WT)	5.5	JUNO (WT)	16,000	-19.9	8.9	$1.0^\dagger$

Results are expressed as means  $\pm$  standard deviation.

$^\dagger N$  was fixed at 1.0.

Extended Data Table 2 | Crystallization and cryoprotectant conditions

IZUMO1*		JUNO		IZUMO1-JUNO complex		
		Form 1	Form 2*	Form 1*	Form 2†	Form 3‡
Protein solution	8.0 mg/ml IZUMO1	20 mg/ml JUNO	8.0 mg/ml JUNO	4.1 mg/ml IZUMO1	5.0 mg/ml IZUMO1	5.0 mg/ml IZUMO1
	10 mM Tris-HCl pH 7.5	10 mM Tris-HCl pH 7.5	10 mM Tris-HCl pH 7.5	4.3 mg/ml JUNO	5.0 mg/ml JUNO	5.0 mg/ml JUNO
	0.15 M NaCl	0.15 M NaCl	0.15 M NaCl	10 mM Tris-HCl pH 7.5	10 mM Tris-HCl pH 7.5	10 mM Tris-HCl pH 7.5
				0.15 M NaCl	0.15 M NaCl	0.15 M NaCl
Reservoir solution	12-15% (w/v) PEG6000	20% (w/v) PEG3350	3% (w/v) PEG4000	8-12% PEG4000	1.0 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	20% PEG4000
	0.1 M sodium citrate pH 5.6	0.24 M malonate pH 7.0	5% 2-propanol	0.2 M Li <sub>2</sub> SO <sub>4</sub>	1.0 M KCl	0.2 M Li <sub>2</sub> SO <sub>4</sub>
			0.1 M HEPES-NaOH pH 7.5	0.1 M MES-NaOH pH 6.0	0.1 M HEPES-NaOH pH 7.0	0.1 M MES-NaOH pH 6.0
Cryoprotectant solution	12% (w/v) PEG6000	20% (w/v) PEG3350	3% (w/v) PEG4000	12% PEG4000	1.0 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	20% PEG4000
	0.1 M sodium citrate pH 5.6	0.24 M malonate pH 7.0	5% 2-propanol	0.12 M Li <sub>2</sub> SO <sub>4</sub>	1.0 M KCl	0.2 M Li <sub>2</sub> SO <sub>4</sub>
	0.09 M NaCl	0.09 M NaCl	0.1 M HEPES-NaOH pH 7.5	0.06 M MES-NaOH pH 6.0	0.1 M HEPES-NaOH pH 7.0	0.1 M MES-NaOH pH 6.0
	25% ethylene glycol	25% ethylene glycol	0.1 M NaCl	0.1 M NaCl	0.09 M NaCl	0.09 M NaCl
			25% glycerol	25% ethylene glycol	25% ethylene glycol	25% ethylene glycol

\*Deglycosylated samples were used for crystallizations.

†Isopropylated IZUMO1 (undeglycosylated) and methylated JUNO (deglycosylated) were used for crystallization.

‡Ethylated IZUMO1 (deglycosylated) and ethylated JUNO (undeglycosylated) were used for crystallization.

Extended Data Table 3 | Data collection and refinement statistics

	IZUMO1*			JUNO		IZUMO1-JUNO			
	Native	K <sub>2</sub> OsO <sub>4</sub>	K <sub>2</sub> PtCl <sub>4</sub>	Form 1	Form 2*	Form 1*	Form 1* (S-SAD)	Form 2†	Form 3‡
<b>Data collection</b>									
Beamline	PF-AR NE3A	PF-AR NE3A	PF-AR NE3A	PF-AR NE3A	PF-AR NE3A	PF-AR NE3A	PF-1A	PF-AR NE3A	PF-AR NE3A
Wavelength (Å)	1.0000	1.1400	1.0723	1.0000	1.0000	1.0000	2.7000	1.0000	1.0000
Space group	<i>P</i> 1	<i>P</i> 1	<i>P</i> 1	<i>I</i> 2 <sub>1</sub> 2 <sub>1</sub>	<i>I</i> 2	<i>C</i> 222 <sub>1</sub>	<i>C</i> 222 <sub>1</sub>	<i>C</i> 2	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions									
<i>a</i> , <i>b</i> , <i>c</i> (Å)	64.9, 75.1, 108.9	64.8, 75.0, 107.9	64.8, 75.0, 108.3	51.9, 81.0, 235.1	96.8, 88.6, 108.1	65.2, 144.8, 141.9	64.2, 144.5, 142.0	145.6, 65.4, 77.2	141.8, 144.9, 64.2
$\alpha$ , $\beta$ , $\gamma$ (°)	77.6, 79.1, 70.7	77.9, 78.9, 71.3	78.0, 79.1, 71.0		90, 96.9, 90			90, 104.2, 90	
Resolution (Å)	50.0–2.10 (2.14–2.10)#	50.0–2.50 (2.54–2.50)	50.0–2.20 (2.24–2.20)	50.0–2.00 (2.03–2.00)	44.3–3.23 (3.49–3.23)	50.0–2.90 (2.95–2.90)	47.3–3.20 (3.42–3.20)	50.0–2.90 (2.95–2.90)	48.3–2.86 (3.01–2.86)
<i>R</i> <sub>sym</sub>	0.052 (0.575)	0.067 (0.552)	0.058 (0.383)	0.110 (0.708)	0.152 (0.552)	0.072 (0.721)	0.196 (1.685)	0.084 (0.720)	0.0071 (1.061)
<i>I</i> / $\sigma$ <i>I</i>	21.6 (1.7)	18.9 (1.3)	17.8 (2.0)	26.6 (1.9)	8.0 (2.5)	34.5 (2.0)	36.2 (5.5)	21.6 (2.4)	28.5 (2.7)
Completeness (%)	97.6 (92.0)	97.8 (95.9)	97.7 (96.7)	99.2 (93.8)	99.4 (99.8)	100.0 (100.0)	99.7 (99.1)	99.9 (99.2)	100.0 (100.0)
Redundancy	3.5 (3.3)	3.5 (3.1)	3.5 (3.3)	12.3 (9.6)	3.4 (3.3)	6.5 (6.5)	196.7 (202.5)	6.6 (5.9)	13.2 (13.7)
<b>Refinement</b>									
Resolution (Å)	50.0–2.10			50.0–2.00	44.3–3.23	50.0–2.90		50.0–2.90	48.3–2.86
No. reflections	101,240			32,072	13,842	14,419		15,006	29,880
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	19.6 / 23.1			19.8 / 21.8	21.7 / 24.2	23.0 / 26.3		22.3 / 25.2	20.4 / 23.8
No. atoms									
Protein	11,096			3,228	6394	3,484		3,476	6,986
Water	537			149	0	5		3	4
B-factors									
Protein	48.8			44.7	72.6	105.7		88.4	86.9
Water	42.2			42.1		74.7		42.9	44.8
R.m.s deviations									
Bond lengths (Å)	0.012			0.010	0.010	0.012		0.011	0.011
Bond angles (°)	1.62			1.49	1.46	1.69		1.60	1.60

Each data set was collected with one crystal. Highest resolution shell is shown in parentheses.

\*Deglycosylated samples were used for crystallizations.

†Isopropylated IZUMO1 (undeglycosylated) and methylated JUNO (deglycosylated) were used for crystallization.

‡Ethylated IZUMO1 (deglycosylated) and ethylated JUNO (undeglycosylated) were used for crystallization.

# Proteome-wide covalent ligand discovery in native biological systems

Keriann M. Backus<sup>1\*</sup>, Bruno E. Correia<sup>1\*</sup>, Kenneth M. Lum<sup>1</sup>, Stefano Forli<sup>2</sup>, Benjamin D. Horning<sup>1</sup>, Gonzalo E. González-Pérez<sup>3</sup>, Sandip Chatterjee<sup>3</sup>, Bryan R. Lanning<sup>1</sup>, John R. Teijaro<sup>4</sup>, Arthur J. Olson<sup>2</sup>, Dennis W. Wolan<sup>3</sup> & Benjamin F. Cravatt<sup>1</sup>

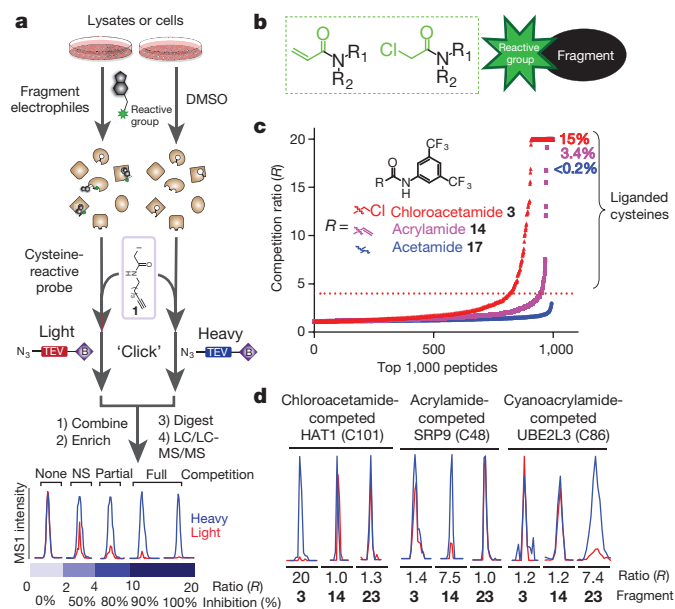
Small molecules are powerful tools for investigating protein function and can serve as leads for new therapeutics. Most human proteins, however, lack small-molecule ligands, and entire protein classes are considered ‘undruggable’<sup>1,2</sup>. Fragment-based ligand discovery can identify small-molecule probes for proteins that have proven difficult to target using high-throughput screening of complex compound libraries<sup>1,3</sup>. Although reversibly binding ligands are commonly pursued, covalent fragments provide an alternative route to small-molecule probes<sup>4–10</sup>, including those that can access regions of proteins that are difficult to target through binding affinity alone<sup>5,10,11</sup>. Here we report a quantitative analysis of cysteine-reactive small-molecule fragments screened against thousands of proteins in human proteomes and cells. Covalent ligands were identified for >700 cysteines found in both druggable proteins and proteins deficient in chemical probes, including transcription factors, adaptor/scaffolding proteins, and uncharacterized proteins. Among the atypical ligand–protein interactions discovered were compounds that react preferentially with pro- (inactive) caspases. We used these ligands to distinguish extrinsic apoptosis pathways in human cell lines versus primary human T cells, showing that the former is largely mediated by caspase-8 while the latter depends on both caspase-8 and -10. Fragment-based covalent ligand discovery provides a greatly expanded portrait of the ligandable proteome and furnishes compounds that can illuminate protein functions in native biological systems.

A major constraint of fragment-based ligand discovery (FBLD) methods is their reliance on assaying purified proteins. This aspect has restricted FBLD to proteins that can be produced in large quantities, and it accordingly remains unclear how many human proteins can be targeted by small molecules or whether these interactions can be optimized to furnish chemical probes for studying protein function in complex biological systems. We aimed to address these questions on a global scale by performing a quantitative analysis of the interactions between fragment electrophiles and thousands of cysteine residues in human proteomes and cells.

We adapted a chemical proteomic method for quantifying cysteine reactivity—termed isotopic tandem orthogonal proteolysis–activity-based protein profiling (isoTOP-ABPP)<sup>12,13</sup>—to perform covalent FBLD in native biological systems. Lysate or intact cells are pre-treated with dimethylsulfoxide (DMSO) or an electrophilic small-molecule fragment and then exposed to a broad-spectrum cysteine-reactive probe, iodoacetamide (IA)-alkyne **1** (Fig. 1a). Proteins harbouring IA-alkyne-labelled cysteine residues from DMSO- and fragment-treated samples are then conjugated by copper-mediated azide-alkyne cycloaddition chemistry<sup>14</sup> to isotopically differentiated azide-biotin tags (heavy and light, respectively), combined, enriched by streptavidin, and proteolytically digested on-bead to yield isotopic peptide pairs that are analysed by liquid chromatography–mass

spectrometry (LC-MS). Quantification of MS1 chromatographic peak ratios for peptide pairs identifies fragment-competed Cys residues as those displaying high competition ratios, or *R* values, in DMSO/fragment comparisons.

We constructed a fragment library predominantly containing chloroacetamide or acrylamide electrophiles (Fig. 1b and Extended Data Fig. 1), which are well-characterized cysteine-reactive groups<sup>10,15–18</sup>. These electrophiles were appended to structurally diverse small-molecule recognition (or binding) elements to create library members



**Figure 1 | Proteome-wide screening of covalent fragments.**

**a**, General protocol for competitive isoTOP-ABPP. Competition ratios, or *R* values, are measured by dividing the MS1 ion peaks for IA-alkyne (1)-labelled peptides in DMSO-treated (heavy, blue) versus fragment-treated (light, red) samples. LC/LC-MS/MS, multidimensional liquid chromatography–tandem mass spectrometry. **b**, General structure of electrophilic fragment library, in which the reactive (electrophilic) and binding groups are coloured green and black, respectively. **c**, Competitive isoTOP-ABPP analysis of the MDA-MB-231 cell proteome pre-treated with the electrophilic 3,5-di(trifluoromethyl)aniline chloroacetamide **3** and acrylamide **14** fragments, along with the non-electrophilic acetamide analogue **17** (500 μM each). Proteomic reactivity values, or liganded cysteine rates, for fragments were calculated as the percentage of total cysteines with *R* values ≥ 4 in DMSO/fragment (heavy/light) comparisons. **d**, Representative MS1 peptide ion chromatograms from competitive isoTOP-ABPP experiments marking liganded cysteines selectively targeted by one of three fragments, **3**, **14** and **23**.

<sup>1</sup>Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA. <sup>2</sup>Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, USA. <sup>3</sup>Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037, USA. <sup>4</sup>Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California 92037, USA.

\*These authors contributed equally to this work.

with an average molecular weight of 284 Da. Since our goal was to probe the ligandability of cysteines in the human proteome, we screened the electrophile library at a high concentration (500  $\mu$ M) similar to the compound concentrations used in FBLD experiments<sup>3</sup>. A subset of the fragment library was initially assayed by competitive profiling in a human MDA-MB-231 breast cancer cell line proteome using IA-rhodamine probe **16**, which permitted SDS–polyacrylamide gel electrophoresis (SDS–PAGE) detection of cysteine reactivity events. This experiment identified several proteins that showed reductions in IA-rhodamine labelling in the presence of one or more fragments (Extended Data Fig. 2a).

We then used competitive isoTOP-ABPP to globally map human proteins and the cysteine residues within these proteins that are targeted by fragment electrophiles. Each fragment was tested against two human cancer cell proteomes (MDA-MB-231 and Ramos cells), and most fragments were screened in duplicate against at least one of these proteomes. On average, 927 cysteines were quantified per data set, and we required that individual cysteines were quantified in at least three data sets for interpretation. On the basis of these criteria, ~6,150 cysteines from ~2,900 proteins were quantified in aggregate across all data sets with an average quantification frequency of 22 data sets per cysteine (Extended Data Fig. 2b). Fragment-competed cysteine residues, or 'liganded' cysteines, were defined as those showing  $\geq 75\%$  reductions in IA-alkyne labelling ( $R$  values  $\geq 4$ ). To minimize the potential for false positives, only cysteines that showed  $R$  values  $\geq 4$  in two or more data sets and met additional criteria for data quality control (see Supplementary Information) were considered as targets of the fragment electrophiles. The proteomic reactivity values, or liganded cysteine rates, of individual fragments were then calculated as the percentage of liganded per total quantified cysteines in isoTOP-ABPP experiments performed on that fragment.

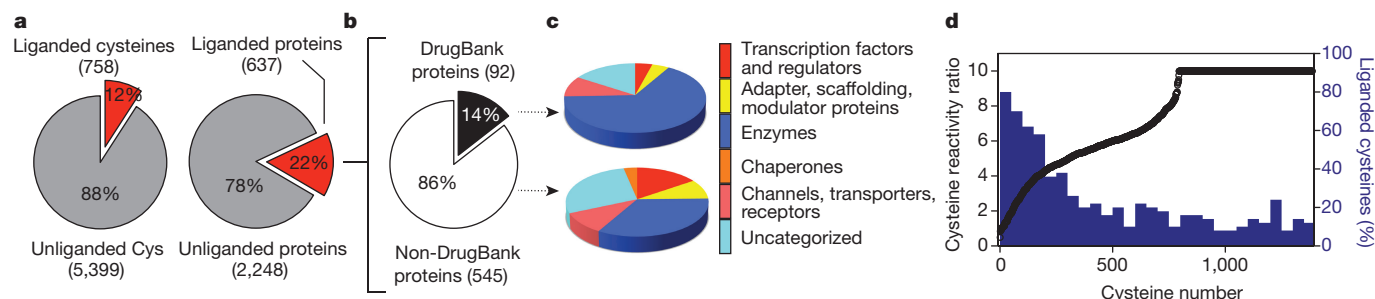
Most fragment electrophiles showed a tempered reactivity across the human proteome, with a median liganded cysteine rate of 3.8% for the library (Extended Data Fig. 2c). Substantial differences in reactivity were observed, however, with individual electrophiles showing liganded cysteine rates of  $<0.1\%$  and others displaying rates  $>15\%$  (Extended Data Fig. 2c). A subset of fragments was also screened at lower concentrations (25–50  $\mu$ M), which confirmed that their proteomic reactivities were concentration-dependent (Extended Data Fig. 2d). The relative reactivity of fragment electrophiles was similar in MDA-MB-231 and Ramos cell proteomes (Extended Data Fig. 2e), indicating that this parameter is an intrinsic property of the compounds. Fragments also showed consistent reactivity profiles when assayed in biological replicate experiments (Extended Data Fig. 2f). We found that the proteomic reactivity of fragment electrophiles was only marginally correlated with their glutathione adduction potential, which is a commonly used surrogate assay for measurements of proteinaceous cysteine reactivity<sup>19</sup> (Extended Data Fig. 2g). We attribute these differences to the impact of

the recognition element of fragment electrophiles on their interactions and reactivity with proteins.

A comparison of fragments **3**, **14**, **17**, and **23–26** provided insights into the relative proteomic reactivity of different electrophilic groups coupled to a common recognition element (3,5-bis(trifluoromethyl) phenyl group). Chloroacetamide **3** exhibited greater reactivity than acrylamide **14** (Fig. 1c), with cyanoacrylamide **23**, but not more sterically congested electrophiles (**24–26**) exhibiting similar reactivity to **14** (Extended Data Fig. 2h). Importantly, the non-electrophilic acetamide control fragment **17** showed negligible activity in competitive isoTOP-ABPP experiments (Fig. 1c), indicating that the vast majority of detected fragment–cysteine interactions reflected covalent reactions versus non-covalent binding events. Also in support of this conclusion, 'clickable' alkyne analogues of **3** and **14** (compounds **19** and **18**, respectively) exhibited different concentration-dependent proteome labelling profiles (**19**  $>$  **18**; Extended Data Fig. 2i) that mirrored the respective liganded cysteine rates of **3** and **14** determined by isoTOP-ABPP (**3**  $>$  **14**; Fig. 1c). Despite the greater overall proteomic reactivity of **3** relative to **14** and **23**, we found clear examples of cysteines that were preferentially liganded by the latter fragments (Fig. 1d and Supplementary Table 1).

Across all isoTOP-ABPP data sets combined, 758 liganded cysteines were identified on 637 distinct proteins, which corresponded to ~12% and 22% of the total quantified cysteines and proteins, respectively (Fig. 2a and Supplementary Table 1). Only a modest fraction of the proteins harbouring liganded cysteines were found in the DrugBank database (14%; Fig. 2b), indicating that the fragment electrophiles targeted many proteins that lack small-molecule probes. Among protein targets with known covalent ligands, the fragment electrophiles frequently targeted the same cysteine residues as these known ligands (Extended Data Table 1a). For one of these targets—the protein kinase BTK—we confirmed that interaction with the drug ibrutinib could be detected by isoTOP-ABPP, which also identified a known ibrutinib off target—MAP2K7 (ref. 20)—in Ramos cell lysates (Extended Data Fig. 3a).

DrugBank proteins with liganded cysteines mostly originated from classes that are regarded as druggable, including enzymes, channels and transporters (Fig. 2c). Non-DrugBank proteins with liganded cysteines, on the other hand, showed a broader class distribution that included proteins, such as transcription factors and adaptor/scaffolding proteins, which are considered challenging to target with small-molecule ligands (Fig. 2c). We previously found that active-site and redox-active cysteines show, in general, greater intrinsic reactivity (as measured with the IA-alkyne probe) compared with other cysteines<sup>12</sup>. While this heightened reactivity appears to be a contributory factor to the ligandability of cysteines, as reflected in the high proportion of hyperreactive (and active-site and redox-active) cysteines discovered as targets of fragment electrophiles, liganded cysteines were also well represented



**Figure 2 | Analysis of cysteines and proteins liganded by fragment electrophiles.** **a**, Fraction of total quantified cysteines and proteins that were liganded by fragment electrophiles in competitive isoTOP-ABPP experiments. **b**, Fraction of liganded proteins found in DrugBank. **c**, Functional classes of DrugBank and non-DrugBank proteins containing liganded cysteines. **d**, Comparison of the ligandability of cysteines as a

function of their intrinsic reactivity with the IA-alkyne probe. Cysteine reactivity values (left y-axis) were taken from ref. 12, where lower ratios correspond to higher cysteine reactivity. A moving average with a step-size of 50 is shown in blue for the percentage of liganded cysteines within each reactivity bin (percentage values shown on right y-axis).

across a broad range of intrinsic reactivities and included many non-active residues (Fig. 2d, Extended Data Fig. 3b, c and Supplementary Discussion). Finally, most proteins were found to harbour a single liganded cysteine among the several cysteines that were, on average, quantified per protein by isoTOP-ABPP (Extended Data Fig. 3d, e and Supplementary Discussion).

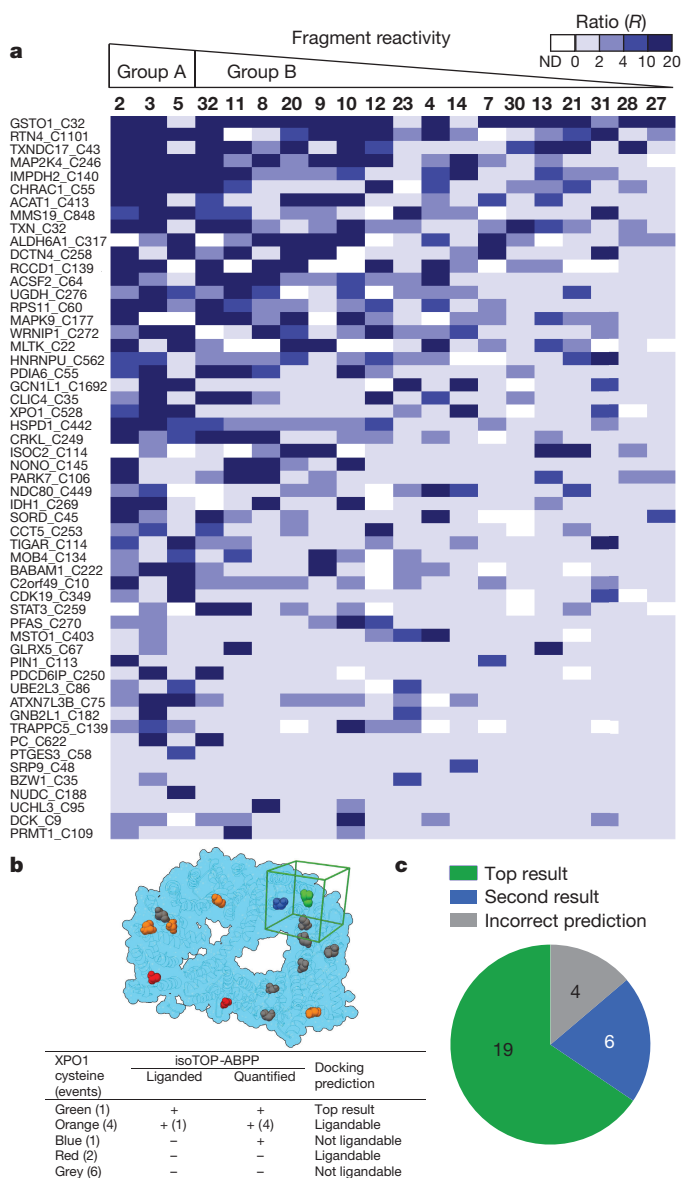
Liganded cysteines, including those found in active and non-active sites of proteins, displayed strikingly distinct structure–activity relationships (SARs) with the fragment electrophile library (Fig. 3a, Extended Data Fig. 3f–l, Supplementary Table 1 and Supplementary Discussion). We also found that, for the majority of liganded cysteines (>60%), electrophile (IA-alkyne or fragment) reactivity was blocked by heat denaturation of the proteome, while only a modest fraction of unliganded cysteines (~20%) showed decreased IA-alkyne labelling after heat denaturation (Extended Data Fig. 3m, n and Supplementary Discussion). These results indicate that the ligand–cysteine interactions are, in general, specific, in that they depend on both the binding groups of ligands and structured sites in proteins (see Supplementary Discussion).

We next asked whether docking could predict sites of fragment electrophile reactivity. Covalent docking programs have recently been introduced to discover ligands that target pre-specified cysteines in proteins<sup>21</sup>; here, however, we aimed to assess computationally the relative ligandability of all cysteines within a protein and match these outputs to the data acquired by isoTOP-ABPP (see Supplementary Discussion). The ranking of our computational predictions matched the experimental data for the majority of proteins investigated (that is, cases in which the top predicted ligandable cysteine matched the liganded cysteine determined by isoTOP-ABPP) (Fig. 3b, c and Extended Data Table 2). We also found that cysteines predicted to be ligandable were much more likely to have been detected by isoTOP-ABPP and exhibit heat-sensitive IA-alkyne reactivity (Extended Data Fig. 3o, p and Extended Data Table 2). These results indicate that reactive docking can provide a good overall prediction of the ligandability of cysteines.

To determine the functional impact of ligand–cysteine interactions mapped by isoTOP-ABPP, we initially selected two enzymes—the protein methyltransferase PRMT1 and the MAP3 kinase MLTK (also known as ZAK)—that possessed liganded cysteines with previously demonstrated activities<sup>12,13</sup>. Our findings confirmed that the fragment electrophiles targeting PRMT1 and MLTK inhibited these enzymes (Extended Data Fig. 4a–d (PRMT1), Extended Data Fig. 4e–i (MLTK) and Supplementary Discussion). We next evaluated proteins that possess previously uncharacterized liganded cysteines, including the nucleotide biosynthetic enzyme IMPDH2 and p53-induced phosphatase TIGAR (p53 also known as TP53). In both cases, we found that ligand–cysteine interactions affected specific functions of these proteins: regulatory nucleotide binding and catalytic activity, respectively (Extended Data Fig. 5a–g (IMPDH2), Extended Data Fig. 5h–n (TIGAR) and Supplementary Discussion).

Competitive isoTOP-ABPP experiments identified distinct subsets of ligands that targeted a conserved cysteine in isocitrate dehydrogenases 1 (IDH1) and 2 (IDH2) (C269 and C308, respectively; Supplementary Table 1). IDH1 and IDH2 are mutated in a number of human cancers to produce enzyme variants with a neomorphic catalytic activity that converts isocitrate to the oncometabolite 2-hydroxyglutarate (2-HG)<sup>22</sup>. Reversible inhibitors selective for mutant forms of IDH1 and IDH2 have been developed and are under clinical investigation for cancer treatment<sup>22</sup>. The liganded cysteine is an active-site-proximal residue that is 13 Å from the NADP<sup>+</sup> molecule in a crystal structure of IDH1 (Extended Data Fig. 6a). We confirmed that fragment ligands inhibited the activity of wild-type but not a C269S mutant of IDH1, and also blocked the R132H oncogenic mutant of IDH1 both *in vitro* and in cells (Extended Data Fig. 6b–k and Supplementary Discussion).

Encouraged by the cellular activity displayed by IDH1 ligands, we sought to more generally assess the capacity of fragment electrophiles to modify cysteines *in situ*. Cells were treated with ~20 representative



**Figure 3 | Analysis of fragment–cysteine interactions.** **a**, Heat map showing R values for representative cysteines and fragments organized by proteomic reactivity values (high to low, left to right) and percentage of fragment hits for individual cysteines (high to low, top to bottom). R values  $\geq 4$  designate fragment hits (coloured medium and dark blue). White indicates not detected (ND). **b**, Representative example of reactive docking predictions shown for XPO1 (Protein Data Bank accession 3GB8). All accessible cysteines were identified and reactive docking was conducted with all fragments from the library within a 25 Å docking cube centred on each accessible cysteine (cube shown in green for liganded Cys in XPO1; see Supplementary Information for more details). Legend presents categories of XPO1 cysteines based on combined docking and isoTOP-ABPP results. **c**, Success rate of reactive docking predictions for liganded cysteines identified by isoTOP-ABPP for 29 representative proteins.

members of the fragment library (50–200  $\mu$ M) for 2 h *in situ* and then harvested, lysed, and analysed by isoTOP-ABPP. The tested fragments showed a broad range of *in situ* reactivities that generally matched their respective reactivities *in vitro*, although some exceptional cases with greater or lesser reactivity *in situ* were noted (Extended Data Fig. 6l and Supplementary Table 1). These differences could reflect the impact of transport and/or metabolic pathways on the cellular concentrations of fragment electrophiles. A substantial fraction (64%) of the liganded cysteines identified in cell lysates were also sensitive to the

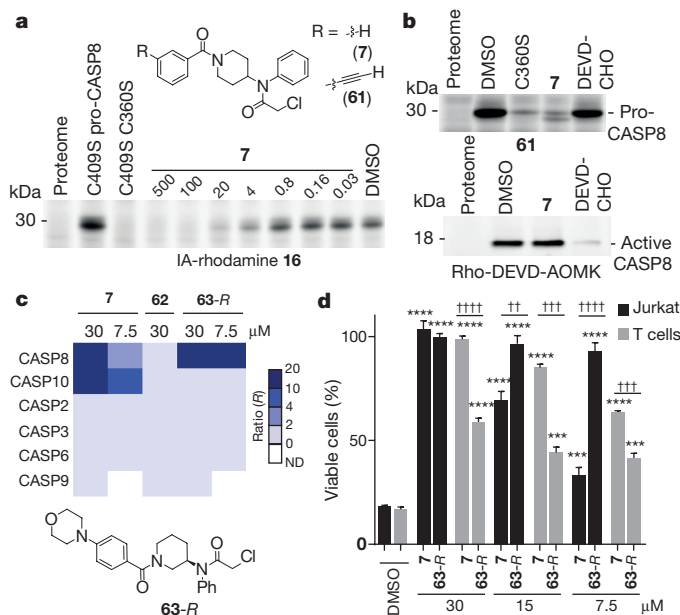
same electrophilic fragments in cells (Extended Data Fig. 6m). Four fragment–cysteine interactions were observed *in situ*, but not in lysates, including C182 of p53, a redox-regulated residue at the dimerization interface of the DNA-binding domain<sup>23</sup> (Extended Data Fig. 6n). These liganded cysteines may require an intact cellular environment to preserve their interactions with fragment electrophiles.

Several fragments targeted the catalytic cysteine nucleophile C360 of the protease caspase-8 (CASP8) in isoTOP-ABPP experiments performed *in vitro* and *in situ* (Extended Data Fig. 7a and Extended Data Table 1). Curiously, however, these fragments exhibited marginal to no inhibition of active CASP8 using either substrate or activity-based probe (Rho-DEVD-AOMK probe) assays (Extended Data Fig. 7b, c). This initially puzzling outcome was explained when we discovered that the electrophilic fragments selectively labelled the inactive zymogen (pro-), but not active form of CASP8 (Fig. 4a, b, Extended Data Fig. 7b–l and Supplementary Discussion). We synthesized a clickable analogue of the most potent pro-CASP8 ligand **7** (**61**; Fig. 4a) and found that this probe (25  $\mu$ M) strongly labelled pro-CASP8, but not a pro-CASP8 C360S mutant (Fig. 4b and Extended Data Fig. 7i), and directly modified C360 of CASP8 in Jurkat cell lysates (Extended Data Table 1b). Compound **7** (50  $\mu$ M) blocked labelling of pro-CASP8 by **61**, but did not inhibit labelling of active CASP8 or other caspases by the Rho-DEVD-AOMK probe<sup>24</sup> (Fig. 4b and Extended Data Fig. 7k, l). Conversely, the general caspase inhibitor Ac-DEVD-CHO (20  $\mu$ M) blocked Rho-DEVD-AOMK labelling of active CASP8 and other caspases, but not **61** labelling of pro-CASP8 (Fig. 4b and Extended Data Fig. 7k, l). Similar results were obtained in substrate assays, in which DEVD-CHO, but not **7**, blocked CASP8 and CASP3 activity (Extended Data Fig. 7b).

We next confirmed that **7**, but not a structurally related inactive probe (**62**; Extended Data Fig. 7f, g, k, m and Supplementary Discussion) blocked Fas ligand (FasL)-, but not staurosporine (STS)-induced apoptosis in Jurkat cells (Extended Data Fig. 7n–p). Chemical proteomic experiments revealed that **7** fully inhibited CASP8, as well as the related initiator caspase CASP10 (but not other caspases, including CASP2, 3, 6 and 9) in Jurkat cells (Fig. 4c, Extended Data Fig. 8a and Supplementary Table 1). We confirmed that **7** blocked labelling of pro-CASP10 by **61** with an apparent half-maximum inhibitory concentration ( $IC_{50}$ ) value of 4.5  $\mu$ M (Extended Data Fig. 8b–d), but did not inhibit active CASP10 as measured by labelling with the Rho-DEVD-AOMK probe (Extended Data Fig. 7l) or a substrate assay (Extended Data Fig. 8e).

The respective functions of CASP8 and CASP10 in extrinsic apoptosis and other cellular processes remain poorly understood<sup>25,26</sup>, in large part due to a lack of selective, non-peptidic, and cell-active inhibitors for these enzymes and the absence of animal models for CASP10 (which is not expressed in rodents). We therefore sought to address this challenge by improving the potency and selectivity of **7**. Conversion of the 4-piperidino moiety to a 3-piperidino group and addition of a *p*-morpholino substituent to the benzoyl ring of **7** furnished compound **63** that was separated by chiral chromatography into its two purified enantiomers, **63-R** (Fig. 4c) and **63-S** (see Supplementary Information), with **63-R** showing substantially improved activity against CASP8 compared to compound **7** (**63-R** apparent  $IC_{50}$  value of 0.7  $\mu$ M (95% confidence interval (CI), 0.5–0.8; Extended Data Fig. 8f–h) and negligible cross-reactivity with CASP10 ( $IC_{50}$  value >100  $\mu$ M; Extended Data Fig. 8c, d, f). **63-S** was much less active against CASP8 (apparent  $IC_{50}$  value of 15  $\mu$ M; Extended Data Fig. 8g, h) and also inactive against CASP10 (Extended Data Fig. 8d). With dual CASP8–CASP10 (**7**) and CASP8-selective (**63-R**) ligands in hand, we next set out to investigate the biological functions of these proteases.

We evaluated the effects of our caspase ligands in human T cells, in which both CASP8 and CASP10 are highly expressed (Extended Data Fig. 8i) and their respective roles much debated<sup>25,26</sup>, as well as in Jurkat cells, which are a commonly studied immortalized human T-cell line. We found that **63-R** fully blocked FasL-induced apoptosis in Jurkat cells and did so with greater potency than **7** (Fig. 4d and



**Figure 4 | Electrophile compounds that target pro-CASP8 and pro-CASP10.** **a**, Compound **7** blocked IA-rhodamine **16** labelling of recombinant, purified pro-CASP8 (bearing a C409S mutation to eliminate **16** labelling at this site; added to Ramos cell lysate at 1  $\mu$ M). Note that a C360S/C409S-mutant of pro-CASP8 did not label with **16**. **b**, **7** blocked probe labelling of pro-, but not active, CASP8. Recombinant pro- and active CASP8 (1  $\mu$ M) were treated with **7** (50  $\mu$ M) or Ac-DEVD-CHO (20  $\mu$ M), for 1 h followed by click probe **61** (25  $\mu$ M) for pro-CASP8 and the Rho-DEVD-AOMK probe (2  $\mu$ M) for active CASP8. **c**, Heat map showing *R* values for caspases measured by quantitative proteomics in Jurkat cells treated with **7**, **63-R** or **62** followed by probe **61** (10  $\mu$ M, 1 h). **d**, Comparison of effects of **7** and **63-R** on FasL-induced apoptosis in Jurkat cells or anti-CD3, anti-CD28-activated primary human T cells. Data represent mean values  $\pm$  standard error of the mean (s.e.m.) for at least three independent experiments, and results are representative of multiple experiments performed with T cells from different human subjects. Statistical significance was calculated with unpaired Student's *t*-tests comparing DMSO- to compound-treated samples, \*\**P* < 0.01, \*\*\*\**P* < 0.0001; and comparing compound **7** to **63-R**, ††*P* < 0.01, †††*P* < 0.001, ††††*P* < 0.0001.

Extended Data Fig. 8j) or **63-S** (Extended Data Fig. 8k). Similar results were obtained in HeLa cells, which express CASP8, but not CASP10 (Extended Data Fig. 8l)<sup>26</sup>. In contrast to these cell line results, FasL-induced apoptosis in primary human T cells showed substantial resistance to **63-R** at all tested concentrations and instead was completely inhibited by the dual CASP8/10 ligand **7** (Fig. 4d). We confirmed by chemical proteomics with probe **61** that **7** blocked both CASP8 and CASP10, while **63-R** inhibited CASP8, but not CASP10, in primary human T cells (Supplementary Table 1) and Jurkat cells (Fig. 4c and Supplementary Table 1). Consistent with these cell death results, **7**, but not **63-R**, prevented proteolytic processing of CASP3 and CASP10 in primary human T cells (Extended Data Fig. 8m). Interestingly, the processing of both CASP8 and the initiator caspase substrate RIP kinase were also preferentially inhibited by **7** versus **63-R** (Extended Data Fig. 8m), indicating that CASP10 may contribute to these proteolytic events in T cells, as has been suggested by biochemical studies<sup>27</sup>. These data, taken together, support substantive functions for both CASP8 and CASP10 in primary human T cells and are consistent with genetic findings indicating that deleterious mutations in either CASP8 or CASP10 can lead to autoimmune syndromes in humans<sup>28</sup>.

By combining chemical proteomics with FBLD, we have found that the human proteome contains many ligandable cysteines. These cysteines were found in proteins not previously known to interact with small molecules, revealing that covalent chemistry can be used to expand the druggable content of the human proteome. Our results for

pro-CASP8 and the results of others for K-Ras(G12C)<sup>10,11</sup> indicate that it is possible to improve the potency and selectivity of covalent fragment hits for protein targets, although the optimization of covalent ligands for cysteines that reside in very shallow pockets may prove more challenging. Some covalent ligands may target cysteines at non-functional sites on proteins, and, in these cases, there is potential to convert the ligands into functional probes using emergent platforms for directing liganded proteins to degradation pathways in the cells<sup>29,30</sup>. We envision that extensions of our chemical proteomic platform could be used to discover ligands that target other nucleophilic amino acids in proteins, thereby increasing the impact covalent chemistry will have on proteome-wide ligand and drug discovery.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 16 January; accepted 13 April 2016.**

**Published online 15 June 2016.**

- Edfeldt, F. N., Folmer, R. H. & Breeze, A. L. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov. Today* **16**, 284–287 (2011).
- Hopkins, A. L. & Groom, C. R. The druggable genome. *Nature Rev. Drug Discov.* **1**, 727–730 (2002).
- Scott, D. E., Coyne, A. G., Hudson, S. A. & Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **51**, 4990–5003 (2012).
- Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proc. Natl Acad. Sci. USA* **97**, 9367–9372 (2000).
- Cardoso, R. *et al.* Identification of Cys255 in HIF-1 $\alpha$  as a novel site for development of covalent inhibitors of HIF-1 $\alpha$ /ARNT PasB domain protein–protein interaction. *Protein Sci.* **21**, 1885–1896 (2012).
- Nonoo, R. H., Armstrong, A. & Mann, D. J. Kinetic template-guided tethering of fragments. *ChemMedChem* **7**, 2082–2086 (2012).
- Kathman, S. G., Xu, Z. & Statsyuk, A. V. A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases. *J. Med. Chem.* **57**, 4969–4974 (2014).
- Jöst, C., Nitsche, C., Scholz, T., Roux, L. & Klein, C. D. Promiscuity and selectivity in covalent enzyme inhibition: a systematic study of electrophilic fragments. *J. Med. Chem.* **57**, 7590–7599 (2014).
- Miller, R. M., Paavilainen, V. O., Krishnan, S., Serafimova, I. M. & Taunton, J. Electrophilic fragment-based design of reversible covalent kinase inhibitors. *J. Am. Chem. Soc.* **135**, 5298–5301 (2013).
- Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–551 (2013).
- Patricelli, M. P. *et al.* Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state. *Cancer Discov.* **6**, 316–329 (2016).
- Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010).
- Wang, C., Weerapana, E., Blewett, M. M. & Cravatt, B. F. A chemoproteomic platform to quantitatively map targets of lipid-derived electrophiles. *Nature Methods* **11**, 79–85 (2014).
- Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem. Int. Edn Engl.* **41**, 2596–2599 (2002).
- Liu, Q. *et al.* Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Biol.* **20**, 146–159 (2013).
- Lim, S. M. *et al.* Therapeutic targeting of oncogenic K-Ras by a covalent catalytic site inhibitor. *Angew. Chem. Int. Ed. Engl.* **53**, 199–204 (2014).
- Hoffstrom, B. G. *et al.* Inhibitors of protein disulfide isomerase suppress apoptosis induced by misfolded proteins. *Nature Chem. Biol.* **6**, 900–906 (2010).
- Li, D. *et al.* BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene* **27**, 4702–4711 (2008).
- Wissner, A. *et al.* Synthesis and structure-activity relationships of 6,7-disubstituted 4-anilinoquinoline-3-carbonitriles. The design of an orally active, irreversible inhibitor of the tyrosine kinase activity of the epidermal growth factor receptor (EGFR) and the human epidermal growth factor receptor-2 (HER-2). *J. Med. Chem.* **46**, 49–63 (2003).
- Lanning, B. R. *et al.* A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nature Chem. Biol.* **10**, 760–767 (2014).
- London, N. *et al.* Covalent docking of large libraries for the discovery of chemical probes. *Nature Chem. Biol.* **10**, 1066–1072 (2014).
- DeLaBarre, B., Hurov, J., Cianchetta, G., Murray, S. & Dang, L. Action at a distance: allostery and the development of drugs to target cancer cell metabolism. *Chem. Biol.* **21**, 1143–1161 (2014).
- Held, J. M. *et al.* Targeted quantitation of site-specific cysteine oxidation in endogenous proteins using a differential alkylation and multiple reaction monitoring mass spectrometry approach. *Mol. Cell. Proteomics* **9**, 1400–1410 (2010).
- Vickers, C. J., González-Páez, G. E. & Wolan, D. W. Selective detection and inhibition of active caspase-3 in cells with optimized peptides. *J. Am. Chem. Soc.* **135**, 12869–12876 (2013).
- Krammer, P. H., Arnold, R. & Lavrik, I. N. Life and death in peripheral T cells. *Nature Rev. Immunol.* **7**, 532–542 (2007).
- Lafont, E. *et al.* Caspase-10-dependent cell death in Fas/CD95 signalling is not abrogated by caspase inhibitor zVAD-fmk. *PLoS ONE* **5**, e13638 (2010).
- Wachmann, K. *et al.* Activation and specificity of human caspase-10. *Biochemistry* **49**, 8307–8315 (2010).
- Bidère, N., Su, H. C. & Lenardo, M. J. Genetic disorders of programmed cell death in the immune system. *Annu. Rev. Immunol.* **24**, 321–352 (2006).
- Winter, G. E. *et al.* Phthalimide conjugation as a strategy for *in vivo* target protein degradation. *Science* **348**, 1376–1381 (2015).
- Lu, J. *et al.* Hijacking the E3 ubiquitin ligase cereblon to efficiently target BRD4. *Chem. Biol.* **22**, 755–763 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by the National Institutes of Health (CA087660 (B.F.C.), GM090294 (B.F.C.), GM108208 (K.M.B.), GM069832 (S.F., A.J.O.)). We thank J. Cisar, K. Mowen, C. Wang, M. Suciu, M. Dix, G. Simon, M. Carrillo, and J. Hulse for experimental assistance, M. Lenardo, L. Zheng and R. Siegel for helpful suggestions, the Marletta and Vogt laboratories at The Scripps Research Institute for sharing instrumentation, and Iterative Threading ASSEMBly Refinement (I-TASSER) for the structural modelling of IMPDH2.

**Author Contributions** B.F.C. and K.M.B. conceived of the project. K.M.B. performed MS experiments and data analysis. B.E.C. wrote software, compiled and analysed MS data. S.F. wrote software and conducted reactive docking. K.M.B. cloned, overexpressed and purified proteins, and conducted inhibition studies *in vitro* and *in situ*. S.C. cloned and purified IDH1. K.M.B., K.M.L., B.D.H. and B.R.L. synthesized compounds. G.E.G.-P. expressed and purified recombinant caspases and TEV protease. D.W.W. provided assistance with the caspase studies. J.R.T. assisted with the T-cell studies. A.J.O. provided technical advice. K.M.B., B.E.C. and B.F.C. designed experiments and analysed data. K.M.B., B.E.C. and B.F.C. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.F.C. (cravatt@scripps.edu) or K.M.B. (kbackus@scripps.edu).

## METHODS

**Preparation of human cancer cell line proteomes.** With the exception of MUM2C cells, which were provided by M. Hendrix, all cell lines were obtained from ATCC, tested negative for mycoplasma contamination, and were used without further authentication, maintaining a low passage number (<10 passages). Cell lines were grown at 37 °C with 5% CO<sub>2</sub>. MDA-MB-231 (ATCC: HTB-26), HeLa (ATCC: CCL-2) and HEK-293T (ATCC: CRL-3216) cells were grown in DMEM supplemented with 10% fetal bovine serum (FBS), penicillin, streptomycin and glutamine. Jurkat A3 (ATCC: CRL-2570), Ramos (ATCC: CRL-1596) and MUM2C cells were grown in RPMI-1640 medium supplemented with 10% FBS, penicillin and streptomycin. For *in vitro* labelling, cells were grown to 100% confluence for MDA-MB-231 cells or until cell density reached 1.5 million cells per ml for Ramos and Jurkat cells. Cells were washed with cold PBS, scraped with cold PBS and cell pellets were isolated by centrifugation (1,400g, 3 min, 4 °C), and stored at –80 °C until use. Cell pellets were lysed by sonication and fractionated (100,000g, 45 min) to yield soluble and membrane fractions, which were then adjusted to a final protein concentration of 1.5 mg ml<sup>–1</sup> for proteomics experiments and 1 mg ml<sup>–1</sup> for gel-based ABPP experiments. The soluble lysate was prepared fresh from frozen pellets directly before each experiment. Protein concentration was determined using the Bio-Rad DC protein assay kit.

**Proteomic sample preparation.** IsoTOP-ABPP, stable isotope labelling by amino acids in cell culture (SILAC) and reductive dimethylation for stable isotope labelling (REDIME) samples were prepared and analysed as has been reported previously<sup>12,31–33</sup>. For details see Supplementary Information.

***In vitro* covalent fragment treatment.** All compounds were made up as 50 mM stock solutions in DMSO and were used at a final concentration of 500 µM. Owing to its low solubility in aqueous medium, fragment 4 was screened at a final concentration of 250 µM. Soluble lysates were adjusted to 1.5 mg ml<sup>–1</sup> and, for each profiling sample, 0.5 ml of lysate was treated with 5 µl of the 50 mM compound stock solution or 5 µl of DMSO.

***In situ* covalent fragment treatment.** For *in situ* labelling, MDA-MB-231 cells were grown to 95% confluence and Ramos cells were grown to 1 million cells per ml. The media in all samples was replaced with fresh media, containing 200 µM of the indicated fragments and the cells were incubated at 37 °C for 2 h, washed with cold PBS, scraped into cold PBS and harvested by centrifugation (see earlier). Fragments 2, 3, 8, 9, 10, 12, 13, 14, 21, 27, 28, 29, 31, 33, 38, 45, 51 and 56 were screened at 200 µM *in situ*. Fragments 4 and 11 were screened at 100 µM *in situ*. Fragments 2, 3, 8 and 20 were tested at 50 µM *in situ*.

**Heat inactivation.** For heat inactivation experiments, 500 µl of MDA-MB-231 soluble proteome was denatured (95 °C, 10 min) and allowed to cool to ambient temperature. The denatured sample and corresponding non-denatured, native proteome (500 µl) were then each labelled with 100 µM iodoacetamide alkylne (IA-alkyne, 5 µl of 10 mM stock in DMSO) and evaluated by isoTOP-ABPP.

**R value calculation and processing.** The ratios of heavy/light for each unique peptide (DMSO/compound treated; isoTOP-ABPP ratios, *R* values) were quantified with in-house CIMAGE software<sup>12,13</sup>, using default parameters (3 MS1 acquisitions per peak and signal to noise threshold set to 2.5). Site-specific engagement of electrophilic fragments was assessed by blockade of IA-alkyne probe labelling. For peptides that showed a ≥95% reduction in MS1 peak area from the fragment-treated proteome (light TEV tag) when compared to the DMSO-treated proteome (heavy TEV tag), a maximal ratio of 20 was assigned. Ratios for unique peptide entries are calculated for each experiment; overlapping peptides with the same modified cysteine (for example, different charge states, MudPIT chromatographic steps or tryptic termini) are grouped together and the median ratio is reported as the final ratio (*R*). The peptide ratios reported by CIMAGE were further filtered to ensure the removal or correction of low-quality ratios in each individual data set. The quality filters applied were the following: removal of half tryptic peptides; for ratios with high standard deviations from the median (90% of the median or above) the lowest ratio was taken instead of the median; removal of peptides with *R* = 20 and only a single MS2 event triggered during the elution of the parent ion; manual annotation of all the peptides with ratios of 20, removing any peptides with low-quality elution profiles that remained after the previous curation steps. Proteome reactivity values for individual fragments were computed as the percentage of the total quantified cysteine-containing peptides with *R* values ≥ 4 (defined as liganded cysteines) for each replicate experiment and the final proteome reactivity value was calculated as the mean for all replicate experiments for each fragment from both MDA-MB-231 and Ramos cellular proteomes. See Supplementary Information for additional details.

**Functional annotation of liganded cysteines.** Custom Python scripts were used to compile functional annotations available in the UniProtKB/Swiss-Prot Protein Knowledge<sup>34</sup> database (release-2012\_11). Relevant UniProt entries were mined for available functional annotations at the residue level, specifically for annotations

regarding enzyme catalytic residues (active sites), disulfides (redox active and structural) and metal-binding sites. Liganded proteins were queried against the Drugbank database (v4.2)<sup>35</sup> and fractionated into DrugBank and non-DrugBank proteins. Functional keywords assigned at the protein level were collected from the UniProt database and the DrugBank and non-DrugBank categories were further classified into protein functional classes. Previously collected cysteine reactivity data<sup>12</sup> were re-processed using ProLuCID as detailed in the Supplementary Information. Cysteines found in both the reactivity and ligandability data sets were sorted on the basis of their reactivity values (lower ratio indicates higher reactivity). The moving average of the percentage of total liganded cysteines within each reactivity bin (step-size 50) was taken. Custom Python scripts were developed to collect relevant NMR and X-ray structures from the RCSB Protein Data Bank (PDB)<sup>36</sup>. For proteins without available PDB structures, sequence alignments, performed with BLAST<sup>37</sup> to proteins deposited in the PDB, were used to identify structural homologues. For annotation of active-site and non-active cysteines, enzymes with structures in the PDB were manually inspected to evaluate the location of the cysteine. Cysteines were considered to reside in enzyme active sites if they were within 10 Å of an active-site ligand or residue(s). Cysteines outside of the 10 Å range were deemed non-active-site residues. Histograms of fragment hit-rates across high-coverage, ligandable cysteines, active-site and non-active site cysteines were calculated from the subset of ligandable cysteines quantified in ten or more separate experiments. For analyses of trends within the whole data, including histograms and heat maps, a cell-line merged data set was used in which data from the MDA-MB-231 experiments were taken first and the Ramos data were used if there were no data from MDA-MB-231 experiments for a particular fragment and cysteine. Heat maps were generated in R (v3.1.3) using the heatmap.2 algorithm. Protein structures were rendered using PyMol<sup>38</sup>.

**Reactive cysteine docking.** *In silico* fragment library containing all chloroacetamide and acrylamide fragments from Extended Data Fig. 1 was prepared using Open Babel library<sup>39</sup> with custom Python scripts. Fragments were modelled in their reactive form (that is, with explicit chloroacetamide and acrylamide warheads). Three-dimensional coordinates were generated from SMILES strings, calculating their protonation state at pH 7.4, and then minimizing them using MMFF94s forcefield (50K iterations steepest descent; 90K conjugate gradient); for chiral molecules with undefined configuration, all enantiomers were generated, resulting in 53 total fragments. For each protein, the UniProtKB accession number was used to filter the PDB<sup>36</sup>. Structures determined by X-ray crystallography were selected, privileging higher sequence coverage and structure resolution (see Extended Data Table 2 for selected PDB accessions). When no human structures were available, the closest homologous organism available was selected (for example, PRMT1: *R. norvegicus*). Protein structures were prepared following the standard AutoDock protocol. Waters, salts and crystallographic additives were removed; AutoDockTools<sup>40</sup> was used to add hydrogens, calculate Gasteiger–Marsili charges and generate PDBQT files. MSMS reduced surface method<sup>41</sup> was used to identify accessible cysteines. The protein volume was scanned using a probe radius of 1.5 Å; residues were considered accessible if they had at least one atom in contact with either external surfaces or internal cavities. The fragment library was docked independently on each accessible cysteine using AutoDock 4.2 (ref. 40). A grid box of 24.4 × 24.4 × 24.4 Å was centred on the geometric centre of the residue; thiol hydrogen was removed from the side chain, which was modelled as flexible during the docking; the rest of the structure was kept rigid. A custom 13–7 interaction potential was defined between the nucleophilic sulfur and the reactive carbon in the ligands. The equilibrium distance ( $r_{eq}$ ) was set to the length of the C–S covalent bond (1.8 Å); the potential well depth ( $\epsilon_{eq}$ ) varied between 1.0 and 0.175 to model to the reactivity of the different ligands. For each fragment, potential well depth was determined by dividing its proteomic reactivity percentage by 20, and the value for iodoacetamide was approximated as the maximum (2.5) for reference. The potential was implemented by modifying the force field table of AutoDock. Fragments were docked with no constraints, generating 100 poses using the default GA settings. For each fragment, the best docking score pose was analysed: if the distance between the nucleophilic sulfur and the reactive carbon was ≤2.0 Å, the cysteine was considered covalently modified. If a residue was alkylated by at least one ligand, it was considered labelled. The docking score (that is, negative binding energy) was calculated based on the estimated interaction energy of each fragment in its docked pose. The docking score of the best alkylating fragment defined the labelling score. The residue with the best labelling score was considered the most probable to be labelled.

**CASP3 and CASP8 *in vitro* activity assays.** CASP3 and 8 assays were conducted with CASP8 activity assay kit (BioVision, K112-100) and CASP3 activity assay kit (Invitrogen, EnzChek Caspase-3 Assay Kit), following the manufacturer's instructions. For further details see Supplementary Information.

**Primary human T-cell isolation and stimulation.** All studies using samples from human volunteers follow protocols approved by The Scripps Research Institute institutional review board (protocol no. IRB-15-6682). Blood from healthy donors was obtained after informed consent and peripheral blood mononuclear cells (PBMCs) were purified over Ficoll–Hypaque gradients (Sigma–Aldrich). T cells were purified via negative selection with magnetic beads (EasySep, STEMCELL). The purified T cells were washed with sterile PBS and resuspended in RPMI-1640 supplemented with FBS, penicillin, streptomycin and glutamine (2 million cells per ml) and 200,000 cells per well were seeded on non-treated tissue culture, 96-well transparent plates that had been coated with anti-CD3 (1:200, BioXcell) and anti-CD28 (1:500, Biolegend) in PBS (100 µl per well). The T cells were removed from stimulus after 3 days and cultured in complete RPMI-1640 supplemented with IL-2 (10 µg ml<sup>-1</sup>, eBioscience) for 3–4 additional days.

**Apoptosis assays in primary human T cells with CASP8 inhibitors.** Primary human T cells were stimulated for 3 days with anti-CD3 and anti-CD28, and the cells were then washed and cultured in complete RPMI with IL-2 (10 µg ml<sup>-1</sup>) for 4 additional days. For western blot analysis, 10 ml of stimulated primary human T cells (1.5 million cells per ml) in RPMI with IL-2 were then treated with the indicated compounds for 1 h before addition of FasL (1 µl of 100 µg µl<sup>-1</sup> stock solution of *MegaFasLigand* in water, final concentration = 10 ng ml<sup>-1</sup>, Adipogen). After 3 h, cells were harvested by centrifugation, washed in PBS and lysed in cell lysis buffer (BioVision, 1067-100) with 1 × complete protease inhibitor (Roche) and 40 µg of each sample were separated by SDS–PAGE on 14% polyacrylamide gels. The gels were transferred to nitrocellulose membranes and were immunoblotted overnight with the indicated antibodies. For measurements of cell viability, in triplicate for each condition, 150,000 cells (100 µl of 1.5 million cells per ml) were plated in 96-well optical-bottom plates. FasL was used at the same concentration indicated earlier with a 30 min pre-incubation with compounds at the indicated concentrations, followed by 4 h with FasL or DMSO. Twenty times compound stock solutions were made in RPMI immediately before use. Cell viability was measured with CellTiter-Glo Luminescent Cell Viability Assay (Promega) and was read on a Biotech Synergy 4 plate reader.

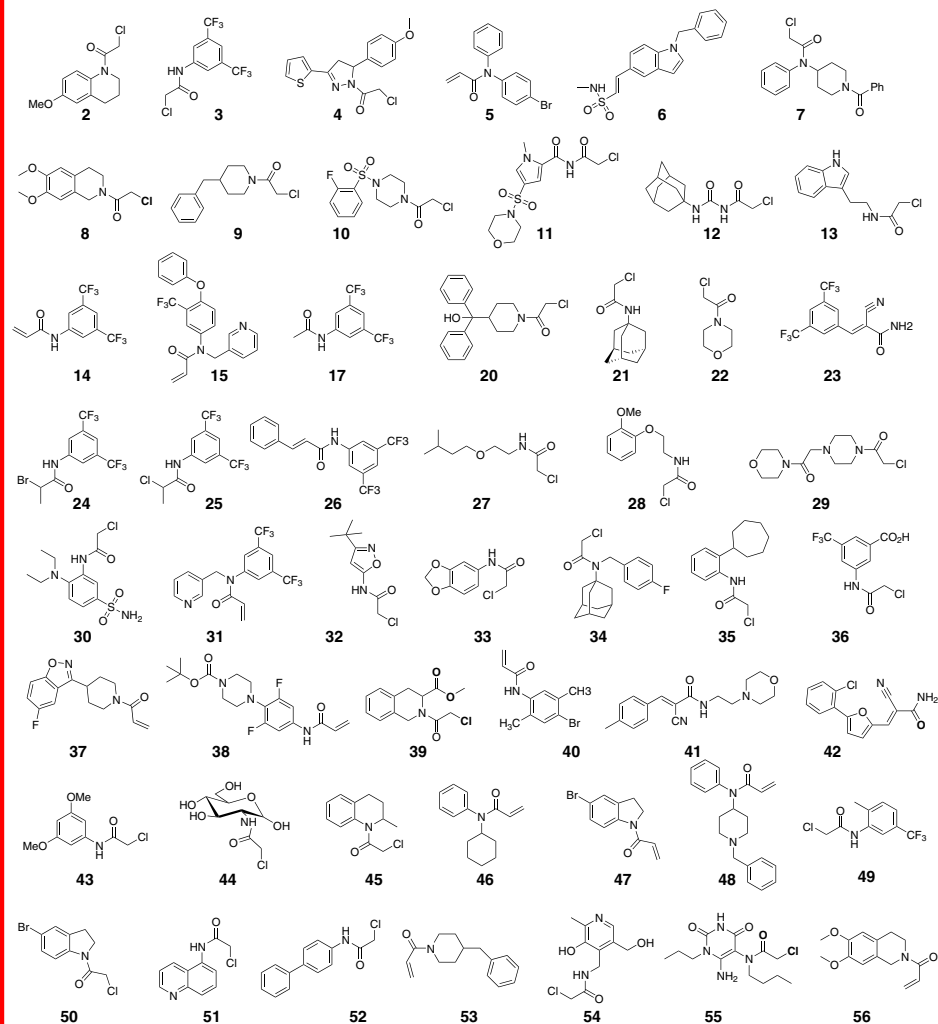
**Western blotting.** For apoptosis studies, cell pellets were resuspended in cell lysis buffer from (BioVision, 1067-100) with 1 × complete protease inhibitor (Roche) and allowed to incubate on ice for 30 min before centrifugation (10 min, 16,000g). For all other studies, cell pellets were resuspended in PBS and lysed with sonication before centrifugation (10 min, 16,000g). The proteins were then resolved by SDS–PAGE and transferred to nitrocellulose membranes, blocked with 5% milk in TBST and probed with the indicated antibodies. The primary antibodies and the dilutions used are as follows: anti-PARP (Cell Signaling, 9532, 1:1,000), anti-CASP3 (Cell Signaling, 9662, 1:1,000), anti-CASP8 (cleaved form, Cell Signaling, 9746, 1:1,000), anti-CASP8 (pro-form, Cell Signaling, 4970, 1:1,000), anti-IDH1

(Cell Signaling, 3997s, 1:500), anti-actin (Cell Signaling, 3700, 1:3,000), anti-GAPDH (Santa Cruz, sc-32233, 1:2,000), anti-Flag (Sigma Aldrich, F1804, 1:3,000), anti-CASP10 (MBL, M059-3, 1:1,000), anti-RIPK (Cell Signaling, 3493S, 1:1,000). Blots were incubated with primary antibodies overnight at 4 °C with rocking and were then washed (3 × 5 min, TBST) and incubated with secondary antibodies (LICOR, IRDye 800CW or IRDye 800LT, 1:10,000) for 1 h at ambient temperature. Blots were further washed (3 × 5 min, TBST) and visualized on a LICOR Odyssey Scanner. The percentage cleavage was determined by quantifying the integrated optical intensity of the bands ( $n = 3$  for STS and  $n = 2$  for FasL), using ImageJ software<sup>42</sup>. For CASP8, the 43 and 41 kDa bands were quantified together. For CASP3, the 17 kDa subunit band was quantified.

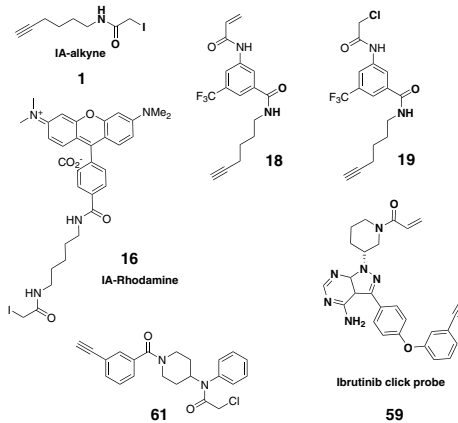
**Statistical analysis.** The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. No statistical methods were used to predetermine sample size. Data are shown as mean ± standard error of the mean. *P* values were calculated using unpaired, two-tailed Student's *t*-test. *P* values of <0.05 were considered significant. See also Supplementary Information.

31. Weerapana, E., Speers, A. E. & Cravatt, B. F. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP)—a general method for mapping sites of probe modification in proteomes. *Nature Protocols* **2**, 1414–1425 (2007).
32. Inloes, J. M. *et al.* The hereditary spastic paraplegia-related enzyme DDHD2 is a principal brain triglyceride lipase. *Proc. Natl Acad. Sci. USA* **111**, 14924–14929 (2014).
33. Adibekian, A. *et al.* Click-generated triazole ureas as ultrapotent *in vivo*-active serine hydrolase inhibitors. *Nat. Chem. Biol.* **7**, 469–478 (2011).
34. Consortium, T. U.; UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
35. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097 (2014).
36. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
37. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
38. DeLano, W. L. *The PyMOL Molecular Graphics System* (Delano Scientific, 2002).
39. O'Boyle, N. M. *et al.* Open Babel: an open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
40. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
41. Sanner, M. F., Olson, A. J. & Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).
42. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
43. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nature Methods* **12**, 7–8 (2015).

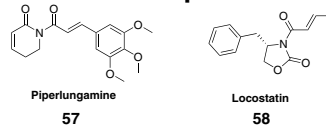
### Fragment electrophiles screened by isoTOP-ABPP



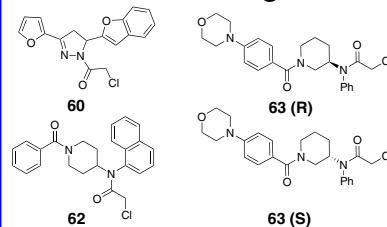
### Click Probes



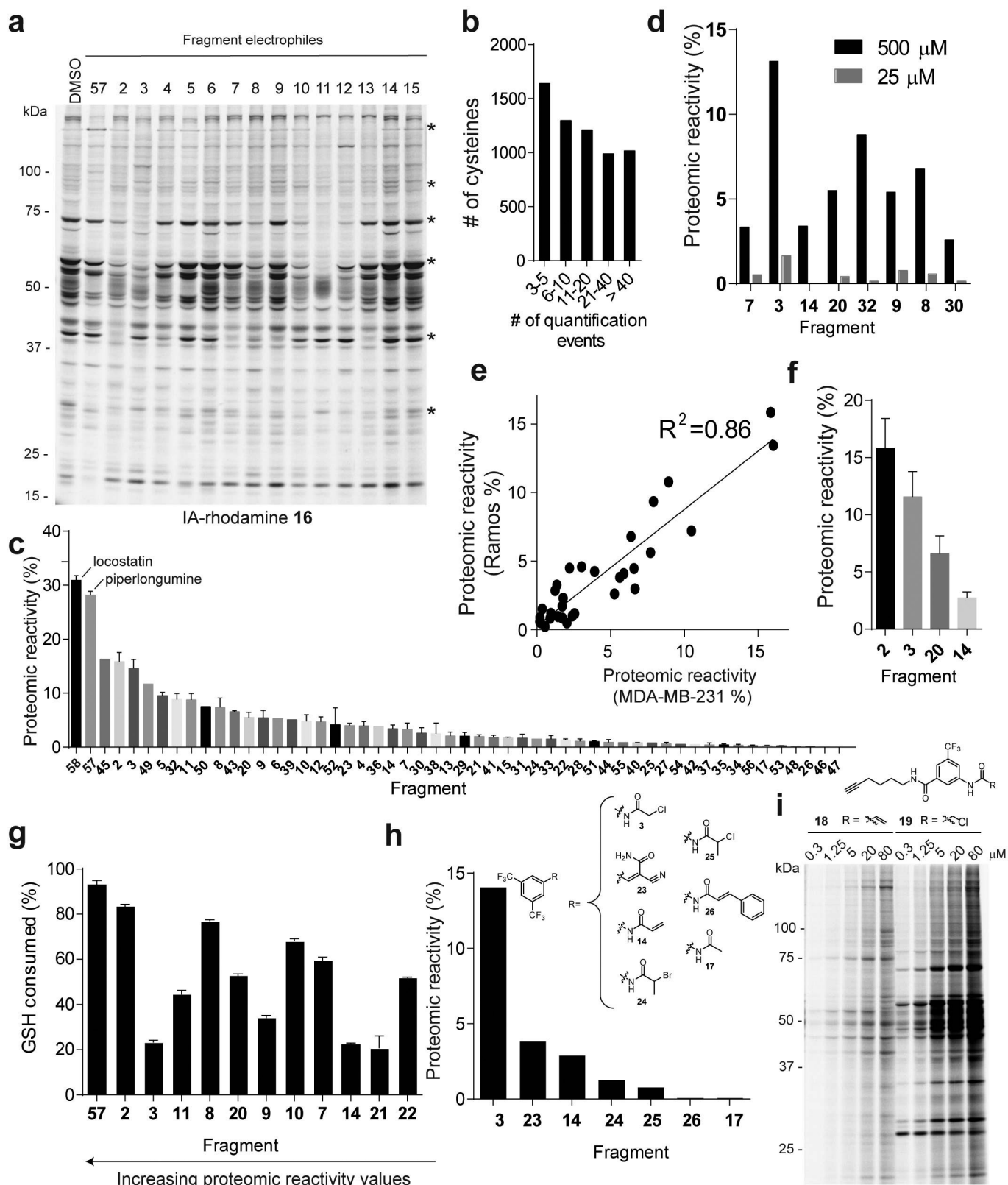
### Tool compounds



### Additional fragments



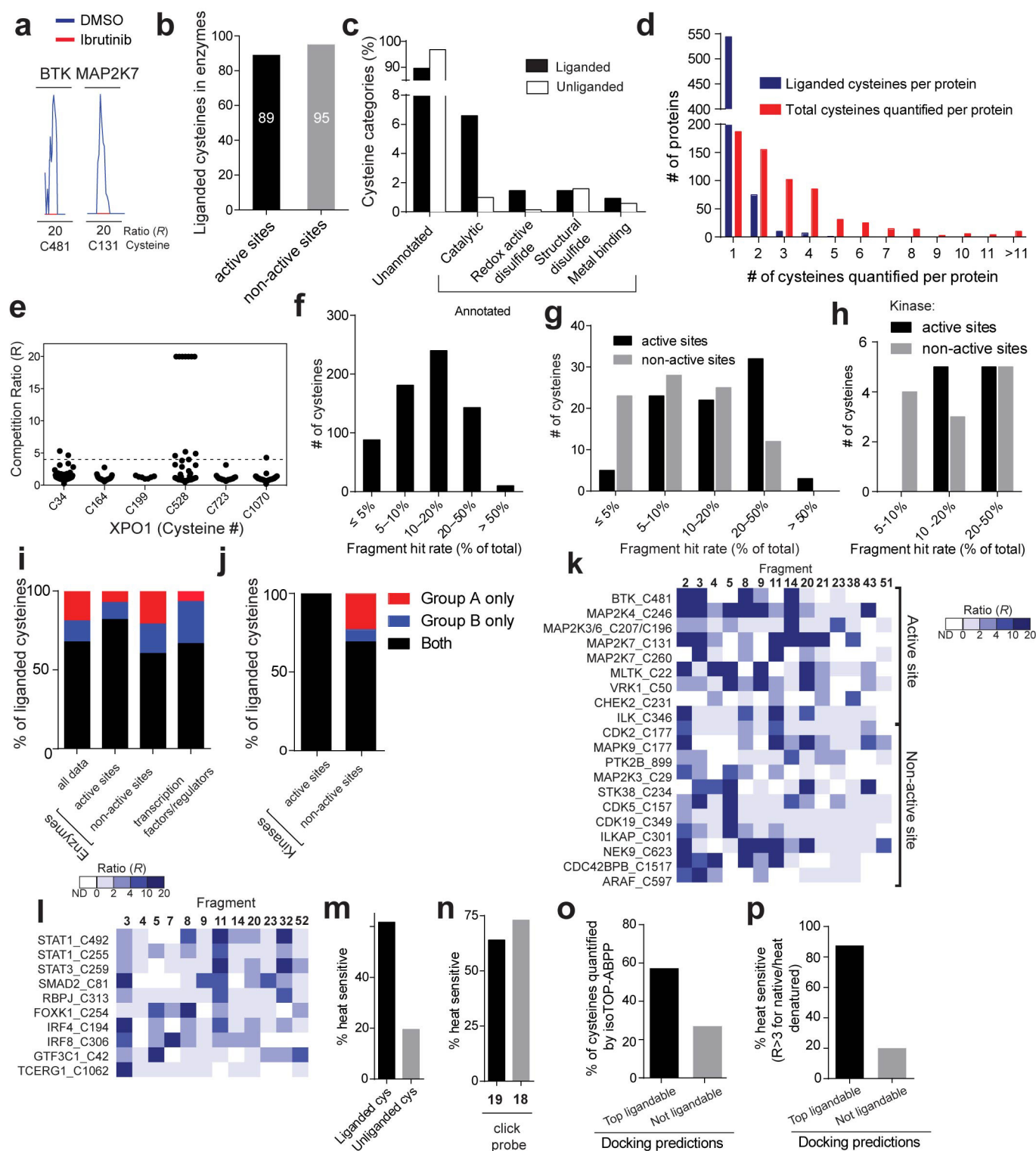
Extended Data Figure 1 | Composition of fragment electrophile library and structures of additional tool compounds, click probes and fragments.



Extended Data Figure 2 | See next page for caption.

**Extended Data Figure 2 | Analysis of proteomic reactivities of fragment electrophiles in human cell lysates.** **a**, Initial analysis of the proteomic reactivity of fragments using an IA-rhodamine probe **16**. Soluble proteome from Ramos cells was treated with the indicated fragments (500  $\mu$ M each) for 1 h, followed by labelling with IA-rhodamine (1  $\mu$ M, 1 h) and analysis by SDS-PAGE and in-gel fluorescence scanning. Several proteins were identified that show impaired reactivity with IA-rhodamine in the presence of one or more fragments (asterisks). Fluorescent gel shown in greyscale. **b**, Frequency of quantification of all cysteines across the complete set of competitive isoTOP-ABPP experiments performed with fragment electrophiles. Note that cysteines were required to have been quantified in at least three isoTOP-ABPP data sets for interpretation. **c**, Rank order of proteomic reactivity values (or liganded cysteine rates) of fragments calculated as the percentage of all quantified cysteines with  $R$  values  $\geq 4$  for each fragment. The majority of fragments were evaluated in 2–4 replicate experiments in MDA-MB-231 and/or Ramos cell lysates,

and their proteomic reactivity values are reported as mean  $\pm$  s.e.m. values for the replicates. **d**, Comparison of the proteomic reactivities of representative fragments screened at 500 versus 25  $\mu$ M in cell lysates. **e**, Comparison of proteomic reactivity values for fragments tested in both Ramos and MDA-MB-231 lysates. **f**, Proteomic reactivity values of representative fragments. **g**, Relative GSH reactivity for representative fragment electrophiles. Consumption of GSH (125  $\mu$ M) was measured using Ellman's reagent (5 mM) after 1 h incubation with the indicated fragments (500  $\mu$ M). **h**, Proteomic reactivity values for fragments electrophiles (500  $\mu$ M) possessing different electrophilic groups attached to a common binding element. **i**, Concentration-dependent labelling of MDA-MB-231 soluble proteomes with acrylamide **18** and chloroacetamide **19** click probes detected by CuACC with a rhodamine-azide tag and analysis by SDS-PAGE and in-gel fluorescence scanning. For **f** and **g**, data represent mean values  $\pm$  s.e.m. for at least three independent experiments.



Extended Data Figure 3 | See next page for caption.

### Extended Data Figure 3 | Analysis of cysteines liganded by fragment electrophiles in competitive isoTOP-ABPP experiments.

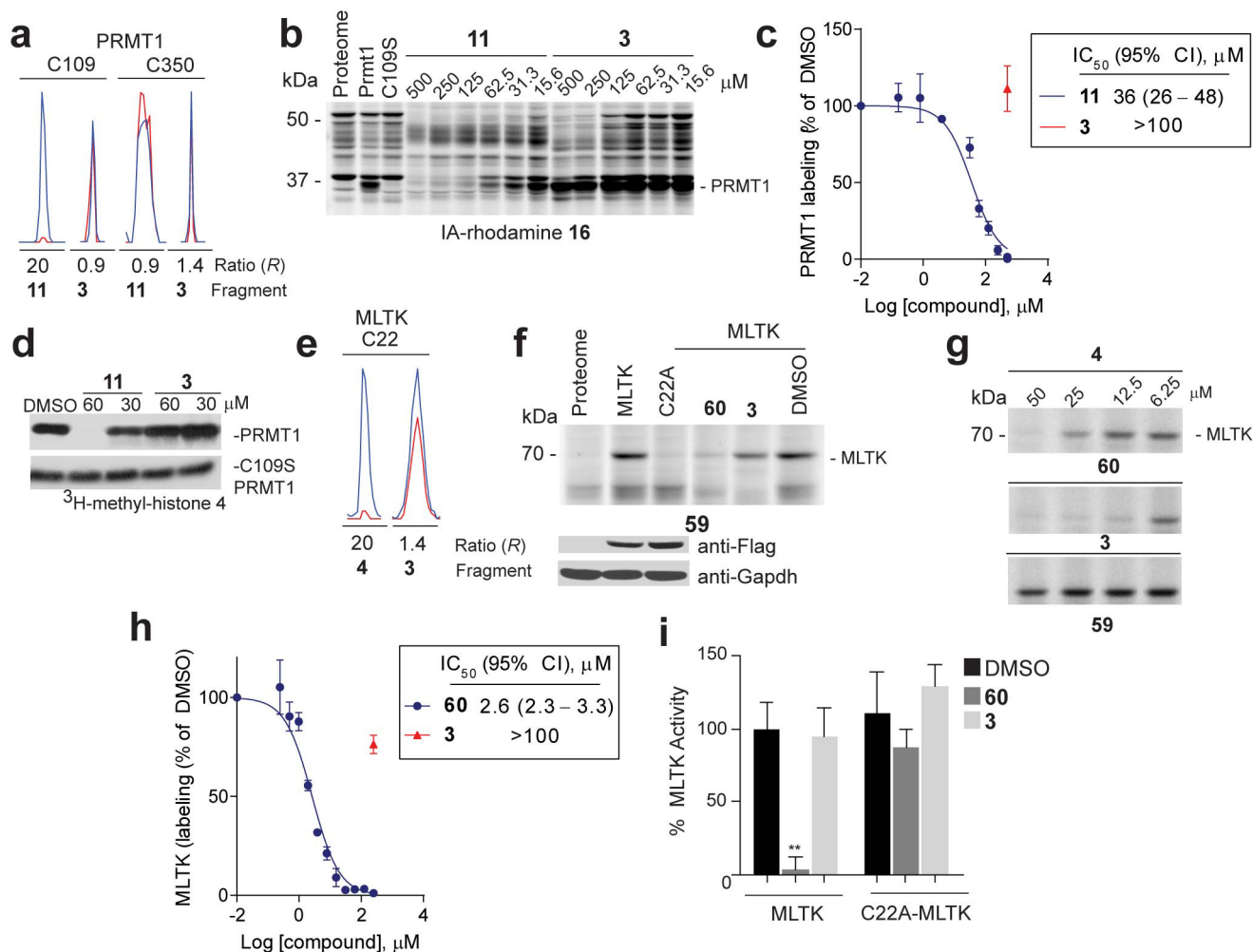
**a**, Representative MS1 ion chromatograms for peptides containing C481 of BTK and C131 of MAP2K7, two cysteines known to be targeted by the anti-cancer drug ibrutinib. Ramos cells were treated with ibrutinib (1  $\mu$ M, 1 h, red trace) or DMSO (blue trace) and evaluated by isoTOP-ABPP.

**b**, Total number of liganded cysteines found in the active sites and non-active sites of enzymes for which X-ray and/or NMR structures have been reported (or reported for a close homologue of the enzyme).

**c**, Functional categorization of liganded and unliganded cysteines based on residue annotations from the UniProt database. **d**, Number of liganded and quantified cysteines per protein measured by isoTOP-ABPP. Respective average values of one and three for liganded and quantified cysteines per protein were measured by isoTOP-ABPP. **e**, *R* values for six cysteines in XPO1 quantified by isoTOP-ABPP, identifying C528 as the most liganded cysteine in this protein. Each point represents a distinct fragment–cysteine interaction quantified by isoTOP-ABPP.

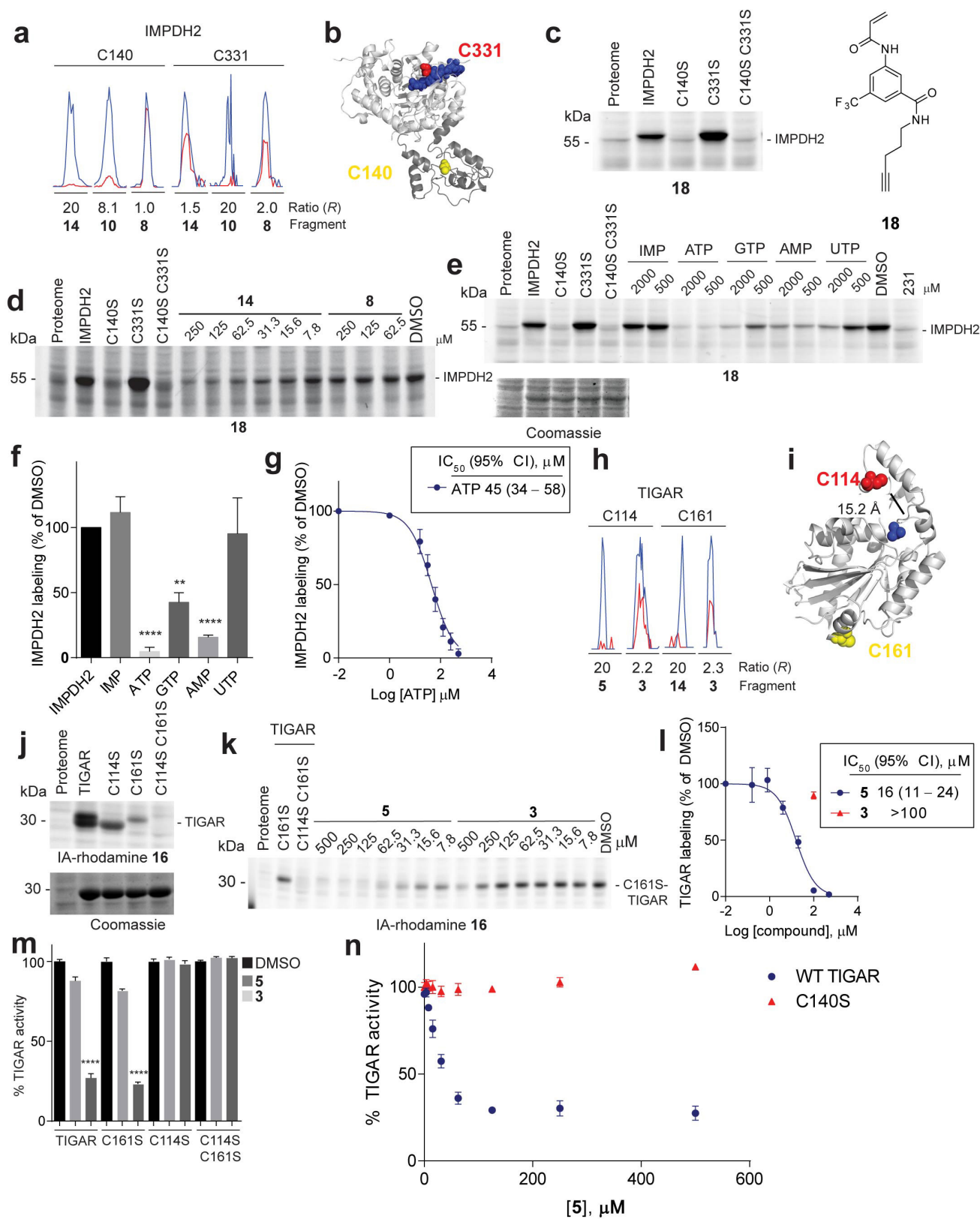
**f–h**, Histograms depicting the percentage of fragments that are hits ( $R \geq 4$ ) for all 768 liganded cysteines (**f**), for liganded cysteines found in enzymes for which X-ray and/or NMR structures have been reported (or reported for a close homologue of the enzyme) (**g**), or for active- and non-active site cysteines in kinases (**h**). **i**, Percentage of liganded cysteines targeted only by group A (red) or B (blue) fragments or both group A and B fragments (black). Shown for all liganded cysteines, liganded cysteines in enzyme

active and non-active sites, and liganded cysteines in transcription factors/regulators. For **g**, **i**, active-site cysteines were defined as those that reside  $<10$  Å from established active-site residues and/or bound substrates/inhibitors in enzyme structures. **j**, The percentage of liganded cysteines in kinases that were targeted by only group A, only group B, or both group A and B compounds. **k**, Heat map showing representative fragment interactions for liganded cysteines found in the active sites and non-active sites of kinases. **l**, Heat map showing representative fragment interactions for liganded cysteines found in transcription factors/regulators. **m**, The fraction of liganded (62%; 341 of 553 quantified cysteines) and unliganded (20%; 561 of 2,870 quantified cysteines) cysteines that are sensitive to heat denaturation measured by IA-alkyne labelling ( $R > 3$  native/heat denatured). **n**, Percentage of proteins identified by isoTOP-ABPP as liganded by fragments **3** and **14** and enriched by their corresponding click probes **19** and **18** that are sensitive to heat denaturation (64% (85 of 133 quantified protein targets) and 73% (19 of 26 quantified protein targets), respectively). Protein enrichment by **18** and **19** was measured by whole-protein capture of isotopically SILAC-labelled MDA-MB-231 cells using quantitative (SILAC) proteomics. **o**, The fraction of cysteines predicted to be ligandable or unligandable by reactive docking that were quantified in isoTOP-ABPP experiments. **p**, The fraction of cysteines predicted to be ligandable or unligandable by reactive docking that show heat-sensitive labelling by the IA-alkyne probe ( $R > 3$  native/heat denatured).



**Extended Data Figure 4 | Confirmation and functional analysis of fragment–cysteine interactions in PRMT1 and MLTK.** **a**, Representative MS1 chromatograms for the indicated Cys-containing peptides from PRMT1 quantified in competitive isoTOP-ABPP experiments of MDA-MB-231 cell lysates, showing blockade of IA-alkyne **1** labelling of C109 by fragment **11**, but not control fragment **3**. **b**, **11**, but not **3**, blocked IA-rhodamine (2  $\mu$ M) labelling of recombinant, purified wild-type PRMT1 (1  $\mu$ M protein doped into HEK293T cell lysates). Note that a C109S PRMT1 mutant did not react with IA-rhodamine. **c**, Apparent IC<sub>50</sub> curve for blockade of **16** labelling of PRMT1 by **11**. CI, 95% confidence intervals. **d**, Effect of **11** and control fragment **3** on methylation of recombinant histone 4 by recombinant PRMT1. Shown is one representative experiment of three independent experiments that yielded similar results. **e**, Representative MS1 ion chromatograms for the MLTK tryptic peptide containing liganded cysteine C22 quantified by isoTOP-ABPP in MDA-MB-231 lysates treated with fragment **4** or control fragment **3** (500  $\mu$ M each). **f**, **60**, but not control fragment **3**

(50  $\mu$ M of each fragment), blocked labelling of recombinant MLTK kinase by a previously reported ibrutinib-derived activity probe **59** (top)<sup>20</sup>. A C22A-MLTK mutant did not react with the ibrutinib probe. Anti-Flag blotting confirmed similar expression of wild-type and C22A-MLTK proteins, which were expressed as Flag-fusion proteins in HEK293T cells (bottom). **g**, Lysates from HEK293T cells expressing wild type or C22A MLTK treated with the indicated fragments and then an ibrutinib-derived activity probe **59** (ref. 20) at 10  $\mu$ M. MLTK labelling by **59** was detected by CuAAC conjugation to a rhodamine-azide tag and analysis by SDS–PAGE and in-gel fluorescence scanning. **h**, Apparent IC<sub>50</sub> curve for blockade of ibrutinib probe-labelling of MLTK by **60**. **i**, **60**, but not control fragment **3** (100  $\mu$ M of each fragment), inhibited the kinase activity of wild-type, but not C22A-MLTK. For **c**, **h** and **i**, data represent mean values  $\pm$  s.e.m. for at least three independent experiments. Statistical significance was calculated with unpaired Student's *t*-tests comparing DMSO- to fragment-treated samples; \*\**P* < 0.1.

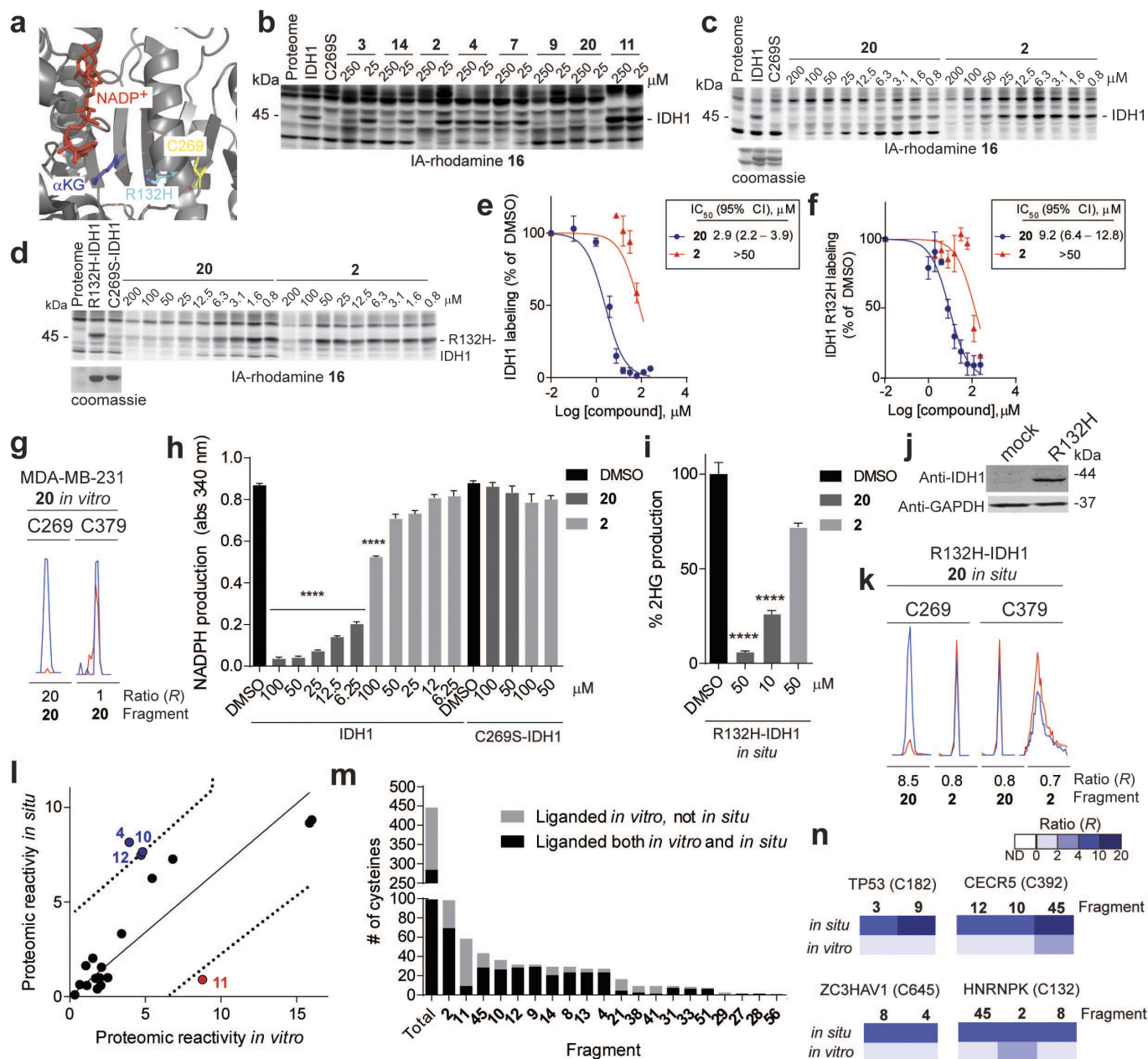


Extended Data Figure 5 | See next page for caption.

### Extended Data Figure 5 | Confirmation and functional analysis of fragment-cysteine interactions in IMPDH2 and TIGAR.

**a**, Representative MS1 ion chromatograms for IMPDH2 tryptic peptides containing the catalytic cysteine, C331, and Bateman domain cysteine, C140, quantified by isoTOP-ABPP in cell lysates treated with the indicated fragments (500  $\mu$ M each). **b**, Structure of human IMPDH2 (PDB accession 1NF7) (light grey) and its structurally unresolved Bateman domain modelled by I-TASSER<sup>43</sup> (dark grey) showing the positions of C331 (red spheres), ribavirin monophosphate and C2-mycophenolic adenine dinucleotide (blue), and C140 (yellow spheres). **c**, Click probe **18** (25  $\mu$ M) labelled wild-type IMPDH2 and C331S IMPDH2, but not C140S IMPDH2 (or C140S/C331S IMPDH2). Labelling was detected by CuAAC conjugation to a rhodamine-azide reporter tag and analysis by SDS-PAGE and in-gel fluorescence scanning. Recombinant IMPDH2 wild type and mutants were expressed and purified from *Escherichia coli* and added to Jurkat lysates to a final concentration of 1  $\mu$ M protein. **d**, Fragment reactivity with recombinant, purified IMPDH2 added to Jurkat lysates to a final concentration of 1  $\mu$ M protein, where reactivity was detected in competition assays using the click probe **18** (25  $\mu$ M). Note that **18** reacted with wild-type and C331S IMPDH2, but not C140S or C140S/C331S IMPDH2. **e**, Nucleotide competition of **18** (25  $\mu$ M) labelling of wild-type IMPDH2 added to MDA-MB-231 lysates to a final concentration of 1  $\mu$ M protein. **f**, Nucleotide competition profile for **18** labelling of recombinant

wild-type IMPDH2 (500  $\mu$ M of each nucleotide). **g**, Apparent IC<sub>50</sub> curve for blockade of **18** labelling of IMPDH2 by ATP. **h**, Representative MS1 chromatograms for TIGAR tryptic peptides containing C114 and C161 quantified by isoTOP-ABPP in cell lysates treated with the indicated fragments (500  $\mu$ M each). **i**, Crystal structure of TIGAR (PDB accession 3DCY) showing C114 (red spheres), C161 (yellow spheres), and inorganic phosphate (blue). **j**, Labelling of recombinant, purified TIGAR and mutant proteins by the IA-rhodamine (2  $\mu$ M) probe. TIGAR proteins were added to MDA-MB-231 lysates, to a final concentration of 2  $\mu$ M protein. **k**, **5**, but not control fragment **3**, blocked IA-rhodamine (2  $\mu$ M) labelling of recombinant, purified C161S TIGAR (2  $\mu$ M protein doped into Ramos cell lysates). **l**, Apparent IC<sub>50</sub> curve for blockade of IA-rhodamine labelling of C161S TIGAR by **5**. **m**, **5**, but not control fragment **3** (100  $\mu$ M of each fragment) inhibited the catalytic activity of wild-type TIGAR, C161S TIGAR, but not C114S TIGAR or C114S/C161S TIGAR. **n**, Concentration-dependent inhibition of wild-type TIGAR by **5**. Note that the C140S-TIGAR mutant was not inhibited by **5**. Data represent mean values  $\pm$  s.e.m. for four replicate experiments at each concentration. For **f**, **g** and **l–n**, data represent mean values  $\pm$  s.e.m. for at least three independent experiments. Statistical significance was calculated with unpaired Student's *t*-tests comparing DMSO- to fragment-treated samples; \*\**P* < 0.01, \*\*\*\**P* < 0.0001.

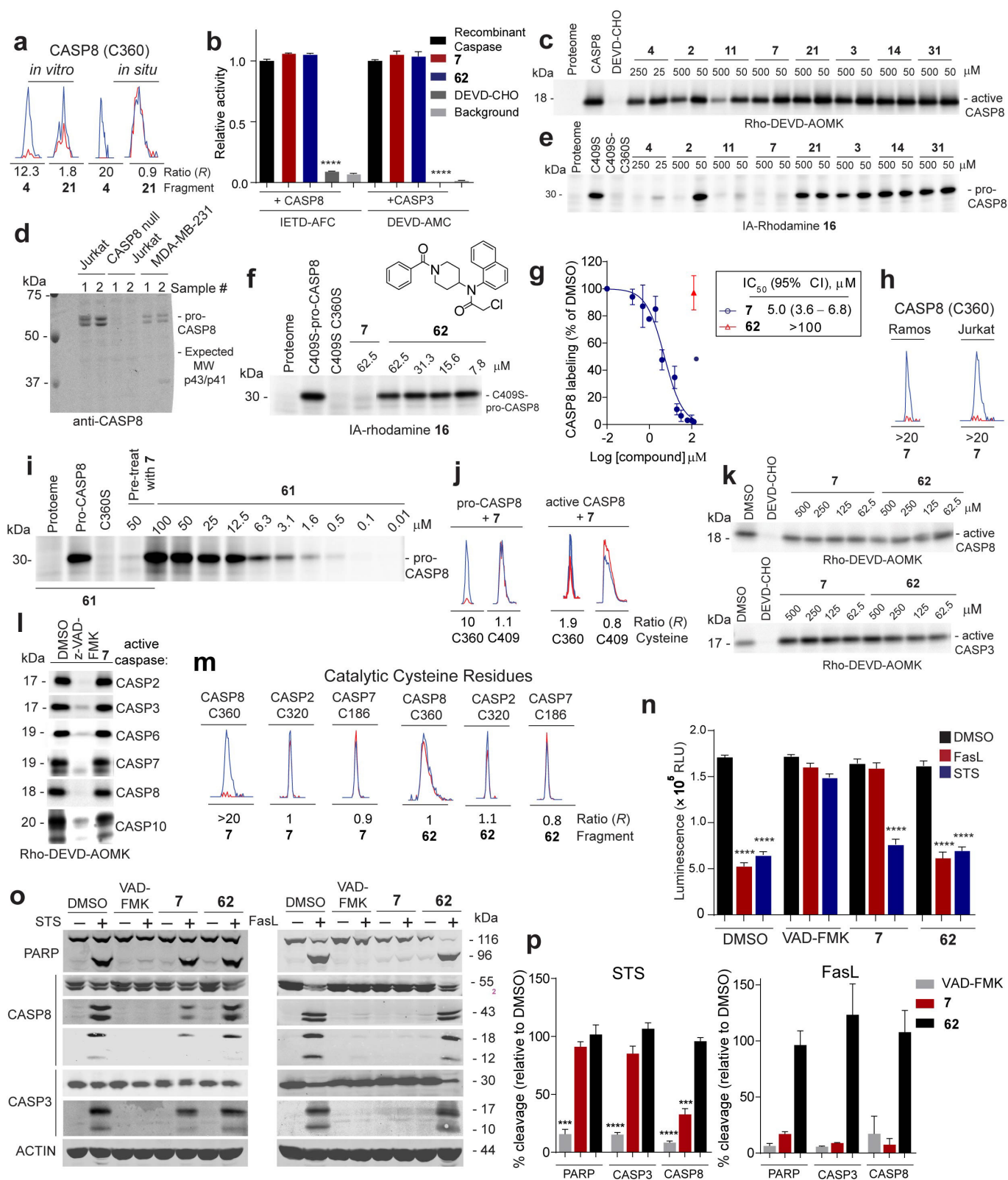


Extended Data Figure 6 | See next page for caption.

**Extended Data Figure 6 | IDH1-related and general *in situ* activity of fragment electrophiles.**

**a**, X-ray crystal structure of IDH1 (PDB accession 3MAS) showing the position of C269 and the frequently mutated residue in cancer, R132. **b**, Blockade of **16** labelling of wild-type IDH1 by representative fragment electrophiles. Recombinant, purified wild-type IDH1 was added to MDA-MB-231 lysates at a final concentration of 2  $\mu$ M, treated with fragments at the indicated concentrations, followed by IA-rhodamine probe **16** (2  $\mu$ M) and analysis by SDS-PAGE and in-gel fluorescence scanning. Note that a C269S mutant of IDH1 did not label with IA-rhodamine **16**. **c**, **d**, Reactivity of **20** and control fragment **2** with recombinant, purified wild-type IDH1 (**b**) or R132H IDH1 (**c**) added to MDA-MB-231 lysates to a final concentration of 2 or 4  $\mu$ M protein, respectively. Fragment reactivity was detected in competition assays using the IA-rhodamine probe (2  $\mu$ M). **e**, **f**, Apparent  $IC_{50}$  curve for blockade of IA-rhodamine labelling of IDH1 (**e**) and R132H IDH1 (**f**) by **20**. Note that the control fragment **2** showed much lower activity. **g**, Representative MS1 ion chromatograms for the IDH1 tryptic peptides containing liganded cysteine C269 and an unliganded cysteine C379 quantified by isoTOP-ABPP in MDA-MB-231 lysates treated with fragment **20** (25  $\mu$ M). **h**, **20**, but not **2**, inhibited IDH1-catalysed oxidation of isocitrate to  $\alpha$ -ketoglutarate ( $\alpha$ -KG) as measured by an increase in NADPH production (340 nm absorbance). **20** did not inhibit the C269S-IDH1 mutant. **i**, **20** inhibited oncometabolite 2-hydroxyglutarate (2-HG) production by R132H IDH1. MUM2C cells stably overexpressing the oncogenic R132H-IDH1 mutant or control green fluorescent protein (GFP)-expressing MUM2C cells were treated with the indicated fragments (2 h, *in situ*). Cells were harvested, lysed and IDH1-dependent production

of 2-HG from  $\alpha$ -KG and NADPH was measured by LC-MS and from which 2-HG production of GFP-expressing MUM2C cells was subtracted (GFP-expressing MUM2C cells produced <10% of the 2-HG generated by R132H-IDH1-expressing MUM2C cells). **j**, Western blot of MUM2C cells stably overexpressing GFP (mock) or R132H-IDH1 proteins. **k**, Representative MS1 chromatograms for the IDH1 tryptic peptides containing liganded cysteine C269 and an unliganded cysteine C379 quantified by isoTOP-ABPP in R132H-IDH-expressing MUM2C lysates treated with **20** or control fragment **2** (50  $\mu$ M, 2 h, *in situ*). **l**, Proteomic reactivity values for individual fragments are comparable *in vitro* and *in situ*. One fragment (**11**) marked in red showed notably lower reactivity *in situ* versus *in vitro*. Reactivity values were calculated as in Fig. 1c. Dashed line mark 90% prediction intervals for the comparison of *in vitro* and *in situ* proteomic reactivity values for fragment electrophiles. Blue and red circles mark fragments that fall above (or just at) or below these prediction intervals, respectively. **m**, Fraction of cysteines liganded *in vitro* that are also liganded *in situ*. Shown are liganded cysteine numbers for individual fragments determined *in vitro* and the fraction of these cysteines that were liganded by the corresponding fragments *in situ*. **n**, Representative cysteines that were selectively targeted by fragments *in situ*, but not *in vitro*. For *in situ*-restricted fragment-cysteine interactions, a second cysteine in the parent protein was detected with an unchanging ratio ( $R \approx 1$ ), thus controlling for potential fragment-induced changes in protein expression. For **e**, **f**, **h** and **i**, data represent mean values  $\pm$  s.e.m. for at least three independent experiments. Statistical significance was calculated with unpaired Student's *t*-tests comparing DMSO- to fragment-treated samples; \*\*\*\* $P < 0.0001$ .



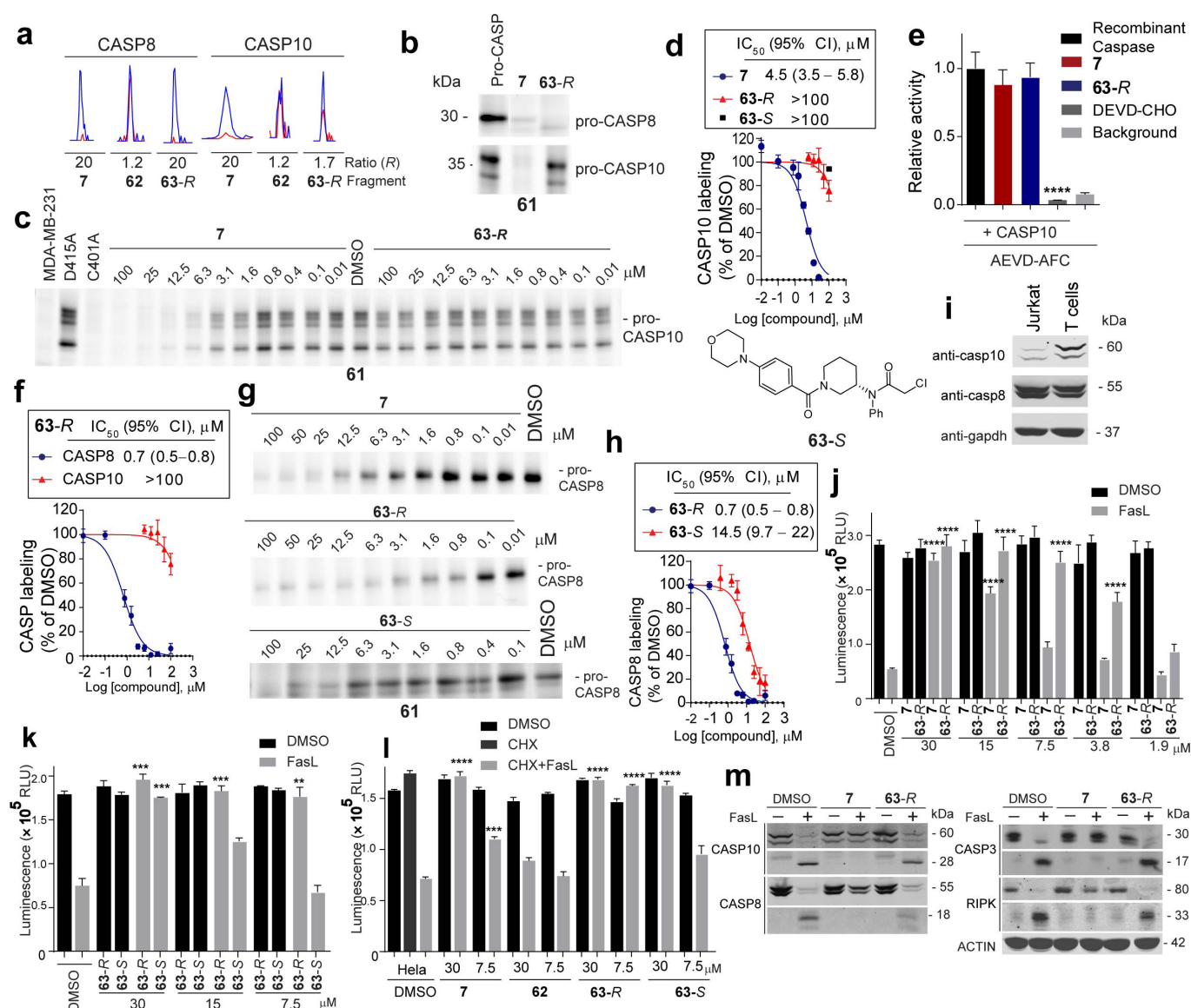
Extended Data Figure 7 | See next page for caption.

**Extended Data Figure 7 | Fragment electrophiles that target**

**pro-CASP8.** **a**, Representative MS1 chromatograms for CASP8 tryptic peptide containing the catalytic cysteine C360 quantified by isoTOP-ABPP in cell lysates or cells treated with fragment **4** (250  $\mu$ M, *in vitro*; 100  $\mu$ M, *in situ*) and control fragment **21** (500  $\mu$ M, *in vitro*; 200  $\mu$ M, *in situ*). **b**, Neither **7** nor control fragment **62** (100  $\mu$ M each) inhibited recombinant, purified active CASP3 and CASP8 assayed using *N*-acetyl-Asp-Glu-Val-Asp-7-amino-4-methylcoumarin (DEVD-AMC) and Ac-Ile-Glu-Thr-Asp-7-amino-4-trifluoromethylcoumarin (IETD-AFC) fluorogenic substrates, respectively. DEVD-CHO (20  $\mu$ M) inhibited both caspases. **c**, Fragment reactivity with recombinant, purified active CASP8 added to cell lysates, where reactivity was detected in competition assays using the caspase activity probe Rho-DEVD-AOMK probe (2  $\mu$ M, 1 h). **d**, Western blot of proteomes from MDA-MB-231, Jurkat, and CASP8-null Jurkat proteomes showing that CASP8 was only found in the pro-enzyme form in these cells. **e**, Fragment reactivity with recombinant, purified pro-CASP8 (D374A, D384A, C409S) added to cell lysates to a final concentration of 1  $\mu$ M protein, where reactivity was detected in competition assays with the IA-rhodamine probe (2  $\mu$ M). Note that mutation of both Cys360 and Cys409 to Ser prevented labelling of pro-CASP8 by IA-rhodamine. **f**, Inactive control fragment **62** did not compete IA-rhodamine labelling of C360 of pro-CASP8. **g**, Apparent IC<sub>50</sub> curve for blockade of IA-rhodamine labelling of pro-CASP8 (C409S) by **7**. **h**, **7** (50  $\mu$ M) fully competed IA-alkyne labelling of C360 of endogenous CASP8 in cell lysates as measured by isoTOP-ABPP. Representative MS1 chromatograms are shown for the C360-containing peptide of CASP8. **i**, Concentration-dependent reactivity of click probe **61**, with recombinant, purified pro-CASP8 (D374A, D384A) added to cell lysates to a final concentration of 1  $\mu$ M protein. Note that pre-treatment with **7** blocked **61** reactivity with pro-CASP8 and mutation of C360 to Ser prevented labelling of pro-CASP8 by **61** (25  $\mu$ M). **j**, **7** (30  $\mu$ M) blocked IA-alkyne labelling of C360 of pro-CASP8, but not active CASP8, as measured by isoTOP-ABPP. Recombinant pro- and active CASP8 were added to Ramos

lysates at 1  $\mu$ M and then treated with **7** (30  $\mu$ M) followed by isoTOP-ABPP. **k**, Fragments **7** and **62** did not block labelling by Rho-DEVD-AOMK (2  $\mu$ M) of recombinant, purified active CASP8 and active CASP3 added to MDA-MB-231 cell lysates to a final concentration of 1  $\mu$ M protein.

**l**, **7** does not inhibit active caspases. Recombinant, active caspases were added to MDA-MB-231 lysate to a final concentration of 200 nM (CASP2, 3, 6, 7) or 1  $\mu$ M (CASP8, 10), treated with *z*-Val-Ala-Asp(OMe)-fluoromethyl ketone (*z*-VAD-FMK) (25  $\mu$ M) or **7** (50  $\mu$ M), followed by labelling with the Rho-DEVD-AOMK probe (2  $\mu$ M). **m**, Representative MS1 chromatograms for tryptic peptides containing the catalytic cysteines of CASP8 (C360), CASP2 (C320), and CASP7 (C186) quantified by isoTOP-ABPP in Jurkat cell lysates treated with **7** or **62** (50  $\mu$ M, 1 h). **n**, **7**, but not control fragment **62**, blocked extrinsic, but not intrinsic apoptosis. Jurkat cells (1.5 million cells per ml) were incubated with **7** or **62** (30  $\mu$ M) or the pan-caspase inhibitor VAD-FMK (100  $\mu$ M) for 30 min before addition of staurosporine (2  $\mu$ M) or *SuperFasLigand* (100 ng ml<sup>-1</sup>). Cells were incubated for 4 h and viability was quantified with CellTiter-Glo. RLU, relative light unit. **o**, For cells treated as described in **n**, cleavage of PARP (96 kDa), CASP8 (p43/p41, p18), and CASP3 (p17) was visualized by western blot. **p**, **7** protects Jurkat cells from extrinsic, but not intrinsic apoptosis. Cleavage of PARP, CASP8 and CASP3 detected by western blotting as shown in **o** was quantified for three (STS) or two (FasL) independent experiments. Cleavage products (PARP (96 kDa), CASP8 (p43/p41), CASP3 (p17)) were quantified for compound treatment and the percentage cleavage relative to DMSO-treated samples was calculated. For **b**, **g** and **n**, data represent mean values  $\pm$  s.e.m. for at least three independent experiments. For **p**, STS data represent mean values  $\pm$  s.e.m. for three independent experiments, and FasL data represent mean values  $\pm$  s.d. for two independent experiments. Statistical significance was calculated with unpaired Student's *t*-tests comparing active compounds (VAD-FMK and **7**) to control compound **62**; \*\**P* < 0.01, \*\*\**P* < 0.001, \*\*\*\**P* < 0.0001.



**Extended Data Figure 8 | CASP10 is involved in intrinsic apoptosis in primary human T cells.** **a**, Representative MS1 peptide signals showing *R* values for caspases detected by quantitative proteomics using probe **61**. ABPP-SILAC experiments. Jurkat cells (10 million cells) were treated with either DMSO (heavy cells) or the indicated compounds (light cells) for 2 h followed by probe **61** (10  $\mu$ M, 1 h). **b**, **7** blocked **61** labelling of pro-CASP8 and CASP10, whereas **63-R** selectively blocked probe labelling of pro-CASP8. **c**, **7**, but not **63-R** blocked probe labelling of pro-CASP10. Recombinant pro-CASP10 was added to MDA-MB-231 lysates to a final concentration of 300 nM, treated with the indicated compounds, and labelled with probe **61**. Mutation of the catalytic cysteine C401A fully prevented labelling by **61**. **d**, Apparent  $IC_{50}$  curve for blockade of **61** labelling of pro-CASP10 by **7**, **63-R** or **63-S**. **e**, Neither **7** nor **63-R** (25  $\mu$ M each) inhibited the activity of recombinant, purified active CASP10 (500 nM), which was assayed after addition of the protein to MDA-MB-231 lysate using fluorometric Ac-Ala-Glu-Val-Asp-7-amino-4-methylcoumarin (AEVD-AMC) substrate. DEVD-CHO (20  $\mu$ M) inhibited the activity of CASP10. **f**, Apparent  $IC_{50}$  curve for blockade of **61** labelling of pro-CASP8 and pro-CASP10 by **63-R**. **g**, **63-R** shows increased potency against pro-CASP8. Recombinant pro-CASP8 was added to MDA-MB-231

lysates to a final concentration of 300 nM, treated with the indicated compounds and labelled with probe **61**. **h**, Apparent  $IC_{50}$  curve for blockade of **61** labelling of pro-CASP8 by **63-R** compared with **63-S**. The structure of **63-S** is shown. **i**, CASP10 is more highly expressed in primary human T cells compared to Jurkat cells. Western blot analysis of full-length CASP10, CASP8 and GAPDH expression levels in Jurkat and T-cell lysates (2 mg ml<sup>-1</sup>). **j**, Jurkat cells (150,000 cells per well) were incubated with **7** or **63-R** at the indicated concentrations for 30 min before addition of staurosporine (2  $\mu$ M) or SuperFasLigand (100 ng ml<sup>-1</sup>). Cells were incubated for 4 h and viability was quantified with CellTiter-Glo (CTG). **k**, Jurkat cells treated as in **j**, but with **63-R** or **63-S**. **l**, HeLa cells (20,000 cells per well) were seeded and 24 h later treated with the indicated compounds for 30 min before the addition of SuperFasLigand (100 ng ml<sup>-1</sup>) and cycloheximide (CHX; 2.5 ng ml<sup>-1</sup>). Cells were incubated for 6 h and viability was quantified with CTG. **m**, For T cells treated as in Fig. 4d cleavage of CASP10 (p22), CASP8 (p18), CASP3 (p17) and RIPK (33 kDa) was visualized by western blotting. For **d-f**, **h** and **j-k**, data represent mean values  $\pm$  s.e.m. for at least three independent experiments. Statistical significance was calculated with unpaired Student's *t*-tests comparing DMSO- to fragment-treated samples; \*\**P* < 0.01, \*\*\*\**P* < 0.0001.

Extended Data Table 1 | Ligandability of representative cysteines and proteins

a.

Protein	Liganded cysteine	Fragment(s)	Other cysteines quantified by isoTOP-ABPP	Previous covalent inhibitor(s)	Cysteine location	Reference
BTK	C481	2, 3, 14, 31	C145, C337	Ibrutinib	Active site	44
TGM2	C277	12, 14, 32	C10, C27, C230, C269, C290, C336, C370, C524, C545, C620	18d	Active Site	45
MAP2K7	C131	2, 3, 11, 14, 20, 21, 38	C260, C280	Ibrutinib	Active Site	20
XPO1	C528	2, 3, 5, 14, 24, 31, 43, 56	C34, C119, C164, C199, C327, C498, C723, C1070	KPT-330	Non-active site	46
CASP5	C315	3, 50	—	Z-WEHD-CHO/FMK	Active Site	47
CASP8	C360	2, 4, 11	C236, C409	Z-VAD-FMK, CV8/9-AOMK	Active Site	48,49
ERCC3	C342	2, 3, 5, 8, 14, 21	—	Triptolide	Active Site	50
PARK7 (Toxoplasma DJ-1)	C106	2, 9, 8, 11, 13, 43, 45, 50, 52	, C46, C53	WRR-086	Active Site	51
GSTO1	C32	2-13, 16, 18-22, 33, 27-30, 32-34, 36, 39, 43, 49, 50, 52, 54, 55	C90, C192, C237	KT53	Active Site	52
ALDH2	C319	3, 8-10, 12, 27, 28, 32, 39, 40, 43, 49, 50	C66, C179, C386, C472	Disulfiram	Active Site	53
CTS2	C92	4, 11, 20, 28, 32	C89, C126, C132, C154, C170, C173, C179, C214	Cy5DCG04	Active Site	54

b.

Protein	Cysteine	Fragment #	Peptide	M+H calculated (m/z)	M+H observed (m/z)	Charge
IMPDH2	C140	14	R.HGFCGIPITDTGR.M	715.86	715.86	2
TIGAR	C114	5	R.EECVPVFTPPGGETLDQVK.M	1123.97	1123.97	2
IDH1	C269	20	K.SEGGFIWACK.N	702.84	702.84	2
CASP8	C360	7	K.VFFIQACQGDNYQK.G	660.98	660.98	3
CASP8*	C360	61-TEV-Tag	K.VFFIQACQGDNYQK.G	1195.58 (light) and 1198.59 (heavy)	1195.58 (light) and 1198.59 (heavy)	2

a. Representative cysteines with known covalent ligands targeted by fragment electrophiles in isoTOP-ABPP experiments. b. Site of fragment labelling for recombinant proteins. Underlines mark the fragment-modified cysteines. \*Measured for endogenous protein by isoTOP-ABPP using probe **61**.

Extended Data Table 2 | Reactive docking results for liganded cysteines

Protein	PDB ID:	Most ligandable cysteine by docking.	Cysteine location	Most ligandable cysteine by isoTOP-ABPP	Match	Heat Sensitive
Aldh2	1O05	C319	Active site	C319	Yes	Yes
BTK	1K2P	C481	Active site	C481	Yes	ND
CASP8	1QTN	C360	Active Site	C360	Yes	Yes
CCNB1	2JGZ	C238	Non-active site	C238	Yes	Yes
CDKN3	1FQ1	C39	Non-active site	C39	Yes	Yes
CLIC4	2AEH	C35	Non-active site	C35	Yes	Yes
DTYMK	1E2G	C163	Non-active site	C163	Yes	No
IDH1	3MAP	C269	Non-active site	C269	Yes	Yes
IMPDH2	1NF7	C331	Active site	C331, C140	Yes	Yes
GLRX5	2WUL	C67	Active site	C67	Yes	No
GSTO1	1EEM	C32	Active site	C32	Yes	Yes
NME3	1ZS6	C158	Non-active site	C158	Yes	Yes
PKM	4JPG	C423	Non-active site	C423	Yes	Yes
SRC	2SRC	C277	Active Site	C277	Yes	ND
TIGAR	3DCY	C114	Non-active site	C114, C161	Yes	Yes
TXNDC	1WOU	C43	Active site	C43	Yes	Yes
UGDH	3ITK	C276	Active site	C276	Yes	Yes
UPP1	3EUF	C162	Non-active site	C162	Yes	Yes
XPO1	3GB8	C528	Non-active site	C528	Yes	Yes
CDK5	1UNG	C157	Non-active site	C269	Second	Yes
EDC3	3D3K	C311	Non-active site	C137, C413, C499	Second	ND
NR2F2	3CJW	C213	Non-active site	C326, C213( <i>in situ</i> )	Second	ND
PDCD6IP	2R02	C231	Non-active site	C90	Second	ND
PRMT1	1ORI	C285	Active site	C109	Second	Yes
UBE2S	1ZDN	C118	Non-active site	C95	Second	ND
FNBP1	2EFL	C145	Non-active site	C70	No	ND
HAT1	2P0W	C120	Non-active site	C101	No	Yes
MAPK9	3NPC	C163	Active site	C177	No	ND
STAT1	1YVL	C543	Non-active site	C492, C255	No	ND

Shows the most ligandable cysteine predicted by reactive docking. Match indicates whether the top cysteine by docking matches that identified by isoTOP-ABPP. Heat sensitive column indicates whether the top cysteine identified by covalent docking is sensitive to heat denaturation. ND, not detected.

# Structural basis of $N^6$ -adenosine methylation by the METTL3–METTL14 complex

Xiang Wang<sup>1\*</sup>, Jing Feng<sup>1\*</sup>, Yuan Xue<sup>1\*</sup>, Zeyuan Guan<sup>1</sup>, Delin Zhang<sup>1</sup>, Zhu Liu<sup>2</sup>, Zhou Gong<sup>3</sup>, Qiang Wang<sup>1</sup>, Jinbo Huang<sup>1</sup>, Chun Tang<sup>2,3</sup>, Tingting Zou<sup>1,4</sup> & Ping Yin<sup>1</sup>

**Chemical modifications of RNA have essential roles in a vast range of cellular processes<sup>1–3</sup>.  $N^6$ -methyladenosine ( $m^6A$ ) is an abundant internal modification in messenger RNA and long non-coding RNA that can be dynamically added and removed by RNA methyltransferases (MTases) and demethylases, respectively<sup>2–5</sup>. An MTase complex comprising methyltransferase-like 3 (METTL3) and methyltransferase-like 14 (METTL14) efficiently catalyses methyl group transfer<sup>6,7</sup>. In contrast to the well-studied DNA MTase<sup>8</sup>, the exact roles of these two RNA MTases in the complex remain to be elucidated. Here we report the crystal structures of the METTL3–METTL14 heterodimer with MTase domains in the ligand-free, S-adenosyl methionine (AdoMet)-bound and S-adenosyl homocysteine (AdoHcy)-bound states, with resolutions of 1.9, 1.71 and 1.61 Å, respectively. Both METTL3 and METTL14 adopt a class I MTase fold and they interact with each other via an extensive hydrogen bonding network, generating a positively charged groove. Notably, AdoMet was observed in only the METTL3 pocket and not in METTL14. Combined with biochemical analysis, these results suggest that in the  $m^6A$  MTase complex, METTL3 primarily functions as the catalytic core, while METTL14 serves as an RNA-binding platform, reminiscent of the target recognition domain of DNA  $N^6$ -adenine MTase<sup>9,10</sup>. This structural information provides an important framework for the functional investigation of  $m^6A$ .**

$N^6$ -methyladenosine is a prevalent RNA modification in species including viruses, bacteria<sup>11</sup>, yeasts<sup>12</sup>, plants<sup>13</sup> and mammals<sup>14,15</sup>. It functions in multiple aspects of developmental regulation<sup>16</sup>, the cell cycle<sup>17</sup>, fate determination<sup>18,19</sup>, and the heat shock stress response<sup>20</sup> by affecting aspects of RNA metabolism such as pre-mRNA processing<sup>21</sup>, translation efficiency<sup>22,23</sup>, transcript stability<sup>24</sup> and miRNA biogenesis<sup>25</sup>. Three distinct classes of protein factor are involved in the function of the  $m^6A$  modification<sup>3,5</sup>: ‘writers’ (adenosine MTases)<sup>6,7,16,26</sup>, ‘erasers’ ( $m^6A$ -demethylating enzymes)<sup>27,28</sup> and ‘readers’ ( $m^6A$ -binding proteins)<sup>24,29</sup>. The writers and erasers reversibly install and remove this modification, respectively, thereby generating a dynamic  $m^6A$  landscape<sup>5</sup>. The readers, known as the YTH domain family<sup>24,29</sup>, bind selectively to the  $m^6A$ -containing sequence and contribute to the determination of RNA fate. Although the erasers and readers have been well characterized, the lack of structures of writers remains a major obstacle to the elucidation of the versatile functions of  $m^6A$ .

In humans, two MTases, METTL3 (also known as MT-A70) and METTL14 participate in this modification as ‘writers’<sup>6,7</sup>. Sequence analysis indicates that both proteins belong to the class I MTase family<sup>30</sup> (Extended Data Fig. 1) and they form a core catalytic complex that is regulated by an additional subunit, Wilms’ tumour 1-associating protein (WTAP)<sup>7,16,26</sup>. Individually, METTL3 and METTL14 exhibit comparable weak MTase activity *in vitro*. However, the

METTL3–METTL14 complex has much higher catalytic activity<sup>6,7</sup>. The mechanism by which the MTases functions synergistically awaits structural investigation.

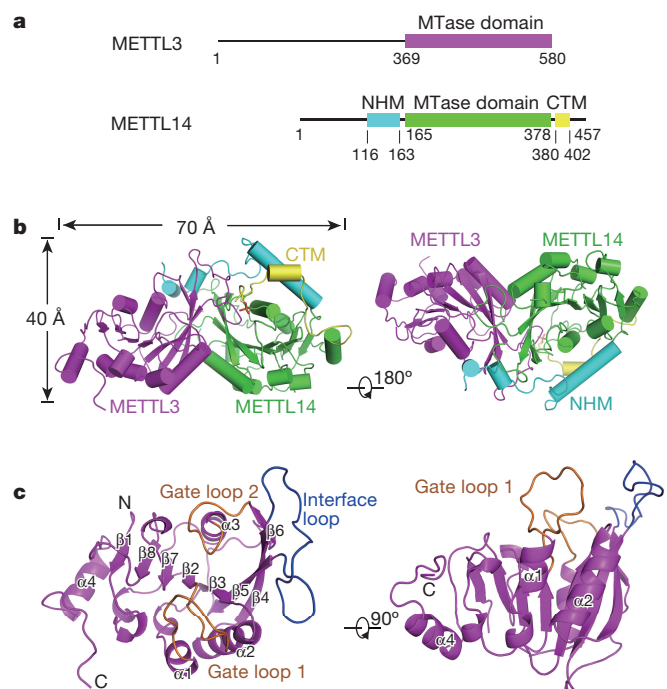
To elucidate the mechanism of  $m^6A$  methylation by the METTL3–METTL14 complex, we determined the crystal structure of the core METTL3–METTL14 complex comprising the MTase domain (METTL3, residues 369–570; METTL14, residues 117–402) (Fig. 1a) in the space group  $P4_12_12$ , using bromide-based single-wavelength anomalous diffraction, at a refined resolution of 1.9 Å (Extended Data Table 1). Additionally, residues 200–204 in METTL14 could not be modelled. In the crystal, one METTL3 molecule and one METTL14 molecule form an antiparallel heterodimer in the asymmetric unit, resulting in an overall butterfly appearance with a width of approximately 40 Å and a length of approximately 70 Å (Fig. 1b).

We traced the METTL3 MTase domain and the METTL14 MTase domain, which is adjacent to the N-terminal  $\alpha$ -helical motif (NHM) and to the C-terminal motif (CTM) with an phosphoserine at position 399 (Fig. 1b). The NHM extends across the MTase domain of METTL14 and to the MTase domain of METTL3. The MTase domain of METTL3 adopts a classic  $\alpha$ – $\beta$ – $\alpha$  sandwich fold comprising a mixed eight-stranded  $\beta$ -sheet with a strand order of  $\beta 1\uparrow$ ,  $\beta 8\uparrow$ ,  $\beta 7\uparrow$ ,  $\beta 2\uparrow$ ,  $\beta 3\uparrow$ ,  $\beta 5\downarrow$ ,  $\beta 4\uparrow$  and  $\beta 6\uparrow$  flanked by four  $\alpha$ -helices ( $\alpha 1$ ,  $\alpha 2$  and  $\alpha 4$  on one side, and  $\alpha 3$  on the other side) and three  $3_{10}$  helices (Fig. 1c). The overall structure of the METTL3 MTase domain primarily resembles that of the class I DNA  $N^6$ -adenine MTase<sup>8,10,30</sup> (Fig. 1c and Extended Data Fig. 2a). Nevertheless, the MTase domains of METTL3 and METTL14 lack an additional element similar to the target recognition domain (TRD) of DNA  $N^6$ -adenine MTase, which functions as the substrate-binding scaffold<sup>8</sup>. Consistent with the 22% sequence identity between the MTase domains of METTL3 and METTL14 (Extended Data Fig. 1), the structures of the two domains were superimposed with a root-mean-squared deviation (r.m.s.d.) of approximately 0.71 Å over 211 C $\alpha$  atoms, excluding the NHM and CTM domains (Extended Data Fig. 2b). Interestingly, three loops with lower sequence similarities exhibited distinct conformations (Extended Data Figs 1, 2b), suggesting that they have different functions. We refer to these loops as gate loop 1 (residues 396–410 in METTL3), interface loop (residues 462–479 in METTL3) and gate loop 2 (residues 507–515 in METTL3; Fig. 1c).

METTL3 and METTL14 adopt the conformation of a tight asymmetric heterodimer with an interface area of approximately 2,500 Å<sup>2</sup> and engage in extensive water-mediated hydrogen-bond interactions with each other. These interactions are mediated by interfaces 1 and 2 (Extended Data Fig. 3a). Interface 1 primarily comprises the interface loop (residues 462–479) of METTL3, NHM, and the long loop connecting  $\beta 5$  and  $\beta 6$  of METTL14 (Extended Data Fig. 3b, left and middle). Additionally, R471 of METTL3 interacts with the C-terminal

<sup>1</sup>National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China. <sup>2</sup>Department of Pharmacology, Zhejiang University, School of Medicine, Hangzhou 310028, China. <sup>3</sup>CAS Key Laboratory of Magnetic Resonance in Biological Systems, State Key Laboratory of Magnetic Resonance and Atomic Molecular Physics, National Center for Magnetic Resonance at Wuhan, Wuhan Institute of Physics and Mathematics of the Chinese Academy of Sciences, Wuhan 430071, China. <sup>4</sup>College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, China.

\*These authors contributed equally to this work.

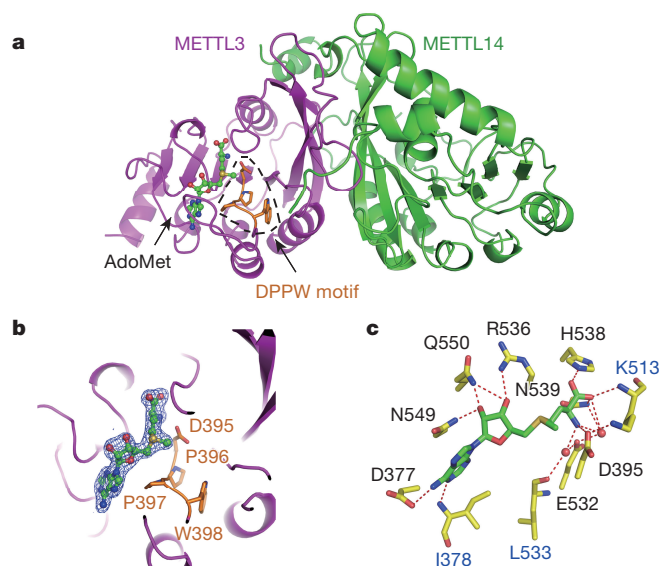


**Figure 1 | Structural overview of the METTL3–METTL14 complex.** **a**, Schematic domain structures of METTL3 and METTL14. METTL3 MTase (residues 369–580), METTL14 NHM (residues 116–163), MTase domain (residues 165–378), and CTM (residues 380–402) are magenta, cyan, green, and yellow, respectively. **b**, Overall structure of the METTL3–METTL14 heterodimer. Residue S399 (red stick) represents the phosphorylation modification. **c**, The METTL3 (magenta) structure in two perpendicular views. Gate loops 1 and 2 are orange, and the interface loop is blue. METTL14 has been removed for clarity. All structure figures were prepared using PyMOL.

phosphorylated S399 of METTL14 via a salt bridge, confirming the important regulatory role of phosphorylation. Interface 2 contains helix  $\alpha 2$  (residues 438–447) and strand  $\beta 4$  (residues 450–460) of METTL3 and the corresponding helix  $\alpha 2$ , strand  $\beta 4$  and an interface loop (residues 266–284) from METTL14 (Extended Data Fig. 3b, right). These interfaces allow the two MTases to bind each other tightly, and the extensive interaction networks are difficult to disrupt.

After extensive trials, we determined the structure of METTL3–METTL14 in complex with AdoMet using a soaking approach and refined the structure to 1.71 Å resolution (Extended Data Table 1). In the crystal, one METTL3–METTL14 heterodimer was present in each asymmetric unit (Fig. 2a). Following assignment of most amino acids into the electron density map, electron densities indicative of one AdoMet became clearly visible in METTL3 (Fig. 2b), whereas no additional electron density was observed in METTL14 (Extended Data Fig. 4). The AdoMet molecule is positioned at the end of  $\beta 1$ ,  $\beta 7$  and  $\beta 8$  (Fig. 2a). The AdoMet-binding site faces the most conserved DPPW motif (residues 395–399) of gate loop 1 (Fig. 2a, b). This orientation suggests a nucleophilic attack methyl-transfer mechanism, reminiscent of the DNA  $N^6$ -adenine MTase<sup>10</sup>.

The AdoMet molecule is primarily coordinated by eleven residues of METTL3 via extensive hydrogen bonds (Fig. 2c). The adenine moiety of AdoMet is recognized by the side chain of D377 and the main chain of I378. The hydroxyl groups of ribose are surrounded by Q550, N549 and R536. Several residues, including D395, K513, H538 and N539, contact the methionine moiety of AdoMet directly, while E532 and L533 form water-mediated interactions with AdoMet (Fig. 2c). The importance of these residues in AdoMet coordination is supported by mutational analysis. D377A, D395A, N539A and E532A mutations completely abolished enzyme activity, while substitutions of R536, H538, N549 or Q550 with alanine moderately reduced enzyme activity



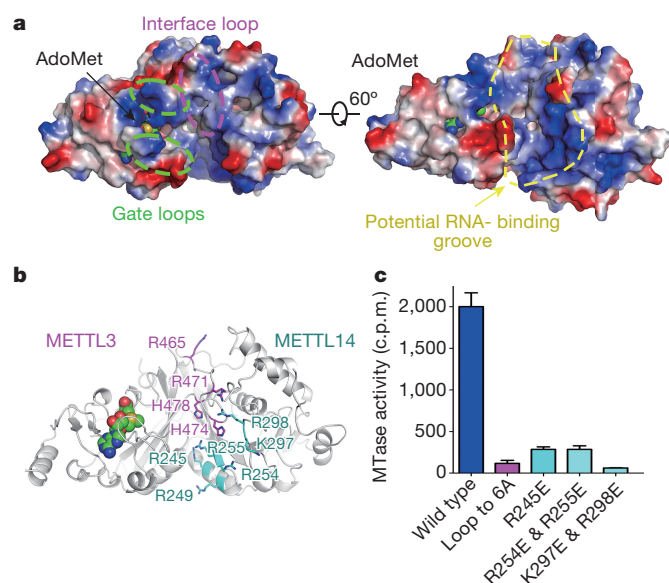
**Figure 2 | AdoMet is coordinated by METTL3 in the binary complex.**

**a**, Ribbon representation of METTL3–METTL14 in complex with AdoMet. The bound AdoMet is illustrated as a green ball-and-stick representation. The DPPW motif is shown as an orange stick surrounded by a dashed line. **b**, Close-up view of the AdoMet binding site and DPPW motif of METTL3 showing the electron density (blue) of AdoMet contoured at  $1\sigma$ . **c**, Schematic representation of the interactions between METTL3 and AdoMet. Residues are shown with sticks. The side-chain interactions and the main-chain interactions are shown in black and blue letters, respectively. Water is shown as a red ball representation. AdoMet is shown as green sticks. Hydrogen bonds are represented as red dashed lines.

(Extended Data Fig. 5a). Neither the D377A nor the D395A mutant had detectable AdoMet-binding activity, as measured by isothermal titration calorimetry (ITC), compared to the wild-type complex, which exhibited a dissociation constant of approximately  $1.5\mu\text{M}$  and binding stoichiometry ( $N$ ) of approximately 1.15 (Extended Data Fig. 5b, c).

Analysis of the surface electrostatic potential of the AdoMet-bound complex revealed a positively charged groove between METTL3 and METTL14 adjacent to the AdoMet (Fig. 3a). This groove consists of at least ten positively charged residues: R465, R471, H474 and H478 from the interface loop of METTL3 and R245, R249, R254, R255, K297 and R298 from METTL14 (Fig. 3b). We speculated that this groove might be responsible for RNA binding. To test this hypothesis, we first replaced the interface loop (residues 462–479) of METTL3 with six alanine amino acids. Compared to the wild-type complex, this mutant exhibited weaker RNA binding activity and reduced MTase activity, but no effect on AdoMet binding was observed (Fig. 3c and Extended Data Fig. 6a, b). A similar result was obtained when the positively charged residues in METTL14 were mutated (Fig. 3c and Extended Data Fig. 6). These results suggest that the positively charged groove formed by METTL3 and METTL14 contributes to internal RNA binding.

The 1.61 Å structure of METTL3–METTL14 in complex with AdoHcy was also determined (Extended Data Table 1). In this structure, one AdoHcy molecule is positioned in the AdoMet-binding pocket of METTL3. AdoHcy adopts a configuration nearly identical to that of AdoMet, except for the ribose (Extended Data Fig. 7a, b). The overall structures of the ligand-free, AdoMet-bound and AdoHcy-bound METTL3–METTL14 complexes are nearly identical, with an r.m.s.d. of 0.24 Å over 438 C $\alpha$  atoms (ligand-free with AdoMet-bound) and 0.12 Å over 454 C $\alpha$  atoms (AdoHcy-bound with AdoMet-bound; Fig. 4a). The structural similarities between the ligand-free and ligand-bound METTL3–METTL14 complexes were also confirmed by small-angle X-ray scattering measurements in solution (Extended Data Fig. 7c).

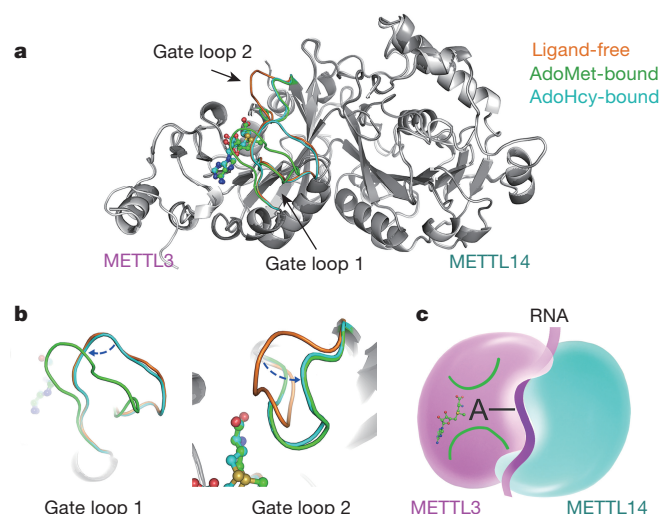


**Figure 3 | Potential RNA-binding groove in the METTL3-METTL14 complex.** **a**, Two views of the surface electrostatic potential calculated with PyMOL. The gate loops and the interface loop of METTL3 are highlighted by green dashed ellipses and a magenta dashed ellipse, respectively (left). The potential RNA-binding groove is encircled by a yellow dashed line (right). **b**, AdoMet in a space-filling representation. Residues from METTL3 and METTL14 are magenta and cyan, respectively. **c**, Measurement of the MTase activity of mutants of the putative RNA-binding groove of the METTL3-METTL14 complex. The indicated mutations were introduced to METTL3 (Loop to 6A) or METTL14 (R245E, R254E & R255E, K297E & R298E). The error bars represent the s.e.m. of three independent measurements.

Close inspection of the structures revealed that gate loop 1 and gate loop 2, which are adjacent to the AdoMet binding site, displayed large conformational changes upon ligand binding (Fig. 4a, b). Gate loop 1 is flipped outwards in the AdoMet-bound state compared with the AdoHcy-bound state, and the ligand-free state adopts the same fold as the AdoHcy-bound state (Fig. 4b). Likewise, gate loop 2 undergoes a significant conformational rearrangement upon binding of AdoHcy or AdoMet, resulting in closure of the binding pocket (Fig. 4a, b). This rearrangement is reminiscent of the interaction of loops 1 and 2 of M.TaqI with its DNA substrate (Extended Data Fig. 8a) and suggests that gate loops 1 and 2 have important roles in adenosine recognition<sup>10</sup>.

The METTL3-METTL14 complex displayed much higher catalytic activity than either METTL3 or METTL14 alone (Extended Data Fig. 5a). This result suggests that METTL3 enhances the MTase activity of METTL14 via RNA binding, vice versa, or both. Our structures suggest that the primary function of METTL14 is not to catalyse methyl-group transfer but to offer an RNA-binding scaffold (Fig. 4c), similar to the TRD of DNA MTases<sup>8–10</sup> (Extended Data Fig. 8a, b). No positively charged area was observed near the potential AdoMet-binding pocket of METTL14 (Extended Data Fig. 8c). Most importantly, although most of the residues involved in AdoMet binding are conserved between METTL3 and METTL14 (Extended Data Figs 1, 8d), mutations of two key residues in METTL14 had little effect on the AdoMet-binding and MTase activities of the binary complex (Extended Data Fig. 8e, f). Consistent with this result, AdoMet was observed in only the METTL3 pocket in the crystal structure (Fig. 2). Nevertheless, we cannot exclude the possibility that METTL14 possesses MTase activity under certain conditions.

A recent study identified a METTL3 protein interaction network comprising additional components such as METTL14, WTAP and KIAA1429 (refs 7, 16, 26). Perturbation of these factors alters global m<sup>6</sup>A levels, resulting in epitranscriptomic changes<sup>4,5</sup>. Additionally,



**Figure 4 | Proposed working model of the METTL3-METTL14 complex.** **a**, Structural superimposition of the ligand-free (orange), AdoMet-bound (green) and AdoHcy-bound (cyan) METTL3-METTL14 complexes shows the conformational changes in gate loops 1 and 2. AdoMet is shown as green balls-and-sticks. **b**, Close-up view of gate loops 1 and 2. The conformational change is highlighted with a blue dashed arrow. **c**, Proposed working model for m<sup>6</sup>A modification by the METTL3-METTL14 complex. METTL3 (magenta) primarily functions as a catalytic core, and METTL14 (cyan) serves as an RNA-binding scaffold. The substrate RNA (magenta ribbon) is cooperatively coordinated by METTL3 and METTL14. The adenine base (black) points to the AdoMet binding site in METTL3 surrounded by the two gate loops (green).

the complex exhibited a substrate sequence preference (Extended Data Fig. 9). Further biochemical and structural characterization of the m<sup>6</sup>A writer complex containing regulatory factors, substrate RNA or both is required to completely elucidate the molecular basis of m<sup>6</sup>A modification. The structures reported here provide unprecedented mechanistic insight into m<sup>6</sup>A RNA methylation and new opportunities for the development of therapeutic agents, and serve as an important foundation for understanding m<sup>6</sup>A epitranscriptomics.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 March; accepted 5 May 2016.**

**Published online 25 May 2016.**

- Lee, M., Kim, B. & Kim, V. N. Emerging roles of RNA modification: m<sup>6</sup>A and U-tail. *Cell* **158**, 980–987 (2014).
- Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N<sup>6</sup>-methyladenosine and gene expression control. *Nature Rev. Mol. Cell Biol.* **15**, 313–326 (2014).
- Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m<sup>6</sup>A RNA methylation. *Nature Rev. Genet.* **15**, 293–306 (2014).
- Schwartz, S. Cracking the epitranscriptome. *RNA* **22**, 169–174 (2016).
- Liu, N. & Pan, T. N-methyladenosine-encoded epitranscriptomics. *Nature Struct. Mol. Biol.* **23**, 98–102 (2016).
- Wang, Y. *et al.* N<sup>6</sup>-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nature Cell Biol.* **16**, 191–198 (2014).
- Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N<sup>6</sup>-adenosine methylation. *Nature Chem. Biol.* **10**, 93–95 (2014).
- Malone, T., Blumenthal, R. M. & Cheng, X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**, 618–632 (1995).
- Gupta, Y. K., Chan, S. H., Xu, S. Y. & Aggarwal, A. K. Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP151. *Nat. Commun.* **6**, 7363 (2015).
- Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J. & Weinhold, E. Structure of the N<sup>6</sup>-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nature Struct. Biol.* **8**, 121–125 (2001).
- Deng, X. *et al.* Widespread occurrence of N<sup>6</sup>-methyladenosine in bacterial mRNA. *Nucleic Acids Res.* **43**, 6557–6567 (2015).

12. Schwartz, S. *et al.* High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409–1421 (2013).
13. Luo, G. Z. *et al.* Unique features of the m<sup>6</sup>A methylome in *Arabidopsis thaliana*. *Nat. Commun.* **5**, 5630 (2014).
14. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
15. Dominissini, D. *et al.* Topology of the human and mouse m<sup>6</sup>A RNA methylomes revealed by m<sup>6</sup>A-seq. *Nature* **485**, 201–206 (2012).
16. Ping, X. L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N<sup>6</sup>-methyladenosine methyltransferase. *Cell Res.* **24**, 177–189 (2014).
17. Fustin, J. M. *et al.* RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell* **155**, 793–806 (2013).
18. Geula, S. *et al.* Stem cells. m<sup>6</sup>A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
19. Chen, T. *et al.* m<sup>6</sup>A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell Stem Cell* **16**, 289–301 (2015).
20. Zhou, J. *et al.* Dynamic m<sup>6</sup>A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
21. Liu, N. *et al.* N<sup>6</sup>-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
22. Wang, X. *et al.* N<sup>6</sup>-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).
23. Choi, J. *et al.* N<sup>6</sup>-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nature Struct. Mol. Biol.* **23**, 110–115 (2016).
24. Wang, X. *et al.* N<sup>6</sup>-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
25. Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N<sup>6</sup>-methyladenosine marks primary microRNAs for processing. *Nature* **519**, 482–485 (2015).
26. Schwartz, S. *et al.* Perturbation of m<sup>6</sup>A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* **8**, 284–296 (2014).
27. Jia, G. *et al.* N<sup>6</sup>-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chem. Biol.* **7**, 885–887 (2011).
28. Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
29. Xiao, W. *et al.* Nuclear m<sup>6</sup>A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* **61**, 507–519 (2016).
30. Iyer, L. M., Zhang, D. & Aravind, L. Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays* **38**, 27–40 (2016).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank B. Sun (SSRF beamline BL17U), R. Zhang (BL19U1), and N. Li (BL19U2) for on-site assistance; S. Fan for data collection support; and research associates at the Center for Protein Research, Huazhong Agricultural University, for technical support. This work was supported by funds from the Ministry of Science and Technology (grants 2015CB910900 and 2013CB900200), Fok Ying-Tong Education Foundation (grant 151021), the Fundamental Research Funds for the Central Universities (Program No. 2014PY026, No. 2015PY219, and No. 2014JQ001), and Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No. 2013RC013).

**Author Contributions** X.W., T.Z. and P.Y. designed all experiments. X.W., J.F. and Y.X. performed protein purification and crystallization. Z.G. determined all of the structures. X.W., Z.L., Z.Go., Q.W., D.Z., J.H., C.T., T.Z. and P.Y. performed the biochemical assays. All authors analysed the data and contributed to manuscript preparation. X.W., T.Z. and P.Y. wrote the manuscript.

**Author Information** The atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank (PDB) with the accession codes 5ILO (ligand-free form), 5IL1 (AdoMet-bound form) and 5IL2 (AdoHcy-bound form). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.Y. ([yingping@mail.hzau.edu.cn](mailto:yingping@mail.hzau.edu.cn)).

**Reviewer Information** *Nature* thanks M. Helm, W. Versées and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**METTL3 and METTL14 construction, expression and purification.** The *METTL3* and *METTL14* genes were amplified from a *Homo sapiens* cDNA library using the following primer pairs: *METTL3*-M1-F (5'-ATGTCGGACACGTGGAGCT-3') and M3ETTL3-L580-R (5'-CTATAAATCTTAGGTCTAGAGATGATAC-3'); and *METTL14*-M1-F (5'-GATAGCCGCTTGCAGGAGATCCG-3') and *METTL14*-R456-R (5'-TCGAGGTGGAAAGCCACCTCTG-3'), and cloned into the T-vector. Both gene strands were verified by sequencing. The full length of genes *METTL3* and *METTL14* was subcloned into a modified pFastBac1 vector with a His<sub>10</sub> affinity tag fused to the N terminus. Bacmids were generated in DH10Bac cells following the instructions for the Bac-to-Bac baculovirus expression system (Invitrogen), and baculoviruses were generated and amplified in Sf-9 insect cells. For protein expression and purification, High Five (*Trichoplusia ni*) insect cells were grown in SIM HF medium (Sino Biological Inc.) supplemented with L-glutamine. The *METTL3*-*METTL14* complex was co-expressed in High Five insect cells at 27°C for 72 h using the *METTL3* and *METTL14* viruses. All complex mutants were co-expressed using a mutant virus and a wild-type partner virus. Cells were harvested by centrifugation at 2,000g for 15 min and homogenized in ice-cold lysis buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl and 0.5 mM phenylmethanesulfonyl-fluoride (PMSF). The cells were disrupted using a cell homogenizer. The insoluble fraction was precipitated by ultracentrifugation (20,000g) for 1 h at 4°C. The supernatant was loaded onto a Ni-NTA superflow affinity column (Qiagen) and washed three times with lysis buffer plus 10 mM imidazole. Elution was performed in buffer containing 25 mM Tris-HCl, pH 8.0, and 250 mM imidazole. The protein was further purified using anion-exchange chromatography (Source 15Q, GE Healthcare). The purified complex was concentrated to approximately 20 mg ml<sup>-1</sup> (Amicon 30-kDa cutoff, Millipore), and if for crystallization, it was digested with chymotrypsin (0.5 mg ml<sup>-1</sup>) at room temperature for 30 min. The undigested or digested protein was subjected to size-exclusion chromatography (Superdex-200 Increase 10/300, GE Healthcare). The buffer used for size-exclusion chromatography contained 25 mM Tris-HCl, pH 8.0, 150 mM NaCl and 5 mM dithiothreitol (DTT). The peak fractions of the *METTL3*-*METTL14* complex were pooled and immediately used for crystallization.

**Crystallization.** Crystallization experiments were performed using the sitting-drop vapour diffusion method at 18°C by mixing equal volumes (1 µl) of protein (15 mg ml<sup>-1</sup>) and reservoir solution. After several rounds of optimization, good-quality crystals appeared after several days and grew as a thin diamond to full size within 15 days in drops containing 18% (v/v) PEG 8000 (Sigma) and 0.1 M sodium citrate, pH 5.7. The crystals were flash-frozen in liquid nitrogen and cryoprotected by adding ethylene glycol to a final concentration of 20%. The crystals were diffracted to 1.9 Å at the Shanghai Synchrotron Radiation Facility on beamlines BL17U and BL19U. To obtain phase information, high-quality crystals were immersed in cryoprotectant solution plus 0.3 M NaBr for 10 min. Before the crystals were harvested, 1 µl of a solution containing 25 mM Tris, pH 8.0, 150 mM NaCl, 20% (v/v) ethylene glycol and 0.1 M NaBr was added. The crystals were immediately transferred to a new solution containing 25 mM Tris, pH 8.0, 150 mM NaCl, ~18% (v/v) ethylene glycol and 0.3 M NaBr. Finally, the bromide-soaked crystals were diffracted to 2.6 Å resolution.

To obtain the AdoMet-bound and AdoHcy-bound structures, we performed extensive trials. We initially failed to obtain diffracting crystals by co-crystallization of *METTL3*-*METTL14* with ligand. We then systematically soaked high-quality diffracting crystals of *METTL3*-*METTL14* with AdoMet or AdoHcy (Sigma). Native crystals of *METTL3*-*METTL14* were obtained after at least 15 days of incubation at 18°C. The crystals were soaked with a series of ligand concentrations in the presence of 3 mM ATP as an additive. The final concentrations of ligand used were 0.5, 1, 2 and 5 mM. The crystals were soaked for 30 min to 72 h, depending on their survival in the soaking solution. The crystals were examined under a microscope every 30 min. If the crystals appeared damaged, they were transferred to a cryoprotectant solution containing 25 mM Tris pH 8.0, 150 mM NaCl and 18% (v/v) ethylene glycol. The crystals were collected and immediately flash-frozen in liquid nitrogen.

**Data collection and structure determination.** All diffraction data were collected at the Shanghai Synchrotron Radiation Facility (SSRF) on beamlines BL17U or BL19U using a CCD detector cooled to 100 K. The data from the *METTL3*-*METTL14* crystals were processed with the HKL2000 program suite and XDS packages<sup>31</sup>. Further processing was performed using the programs from the CCP4 suite<sup>32</sup>. The ligand-free *METTL3*-*METTL14* structure was solved via single anomalous diffraction (SAD) of bromide using the ShelxC/D/E program<sup>33</sup>. Then, a crude model was manually built in the Coot program<sup>34</sup>. The P4<sub>2</sub>2 crystal forms of the

AdoMet-bound and AdoHcy-bound complexes were solved by molecular replacement with PHASER using the structure of the ligand-free *METTL3*-*METTL14* complex as the initial searching model<sup>35</sup>. All four crystal structures were built using Coot and refined using the Phenix program<sup>36</sup>. The data collection and structure refinement statistics are summarized in Extended Data Table 1. All figures representing structures were prepared with PyMOL<sup>37</sup>.

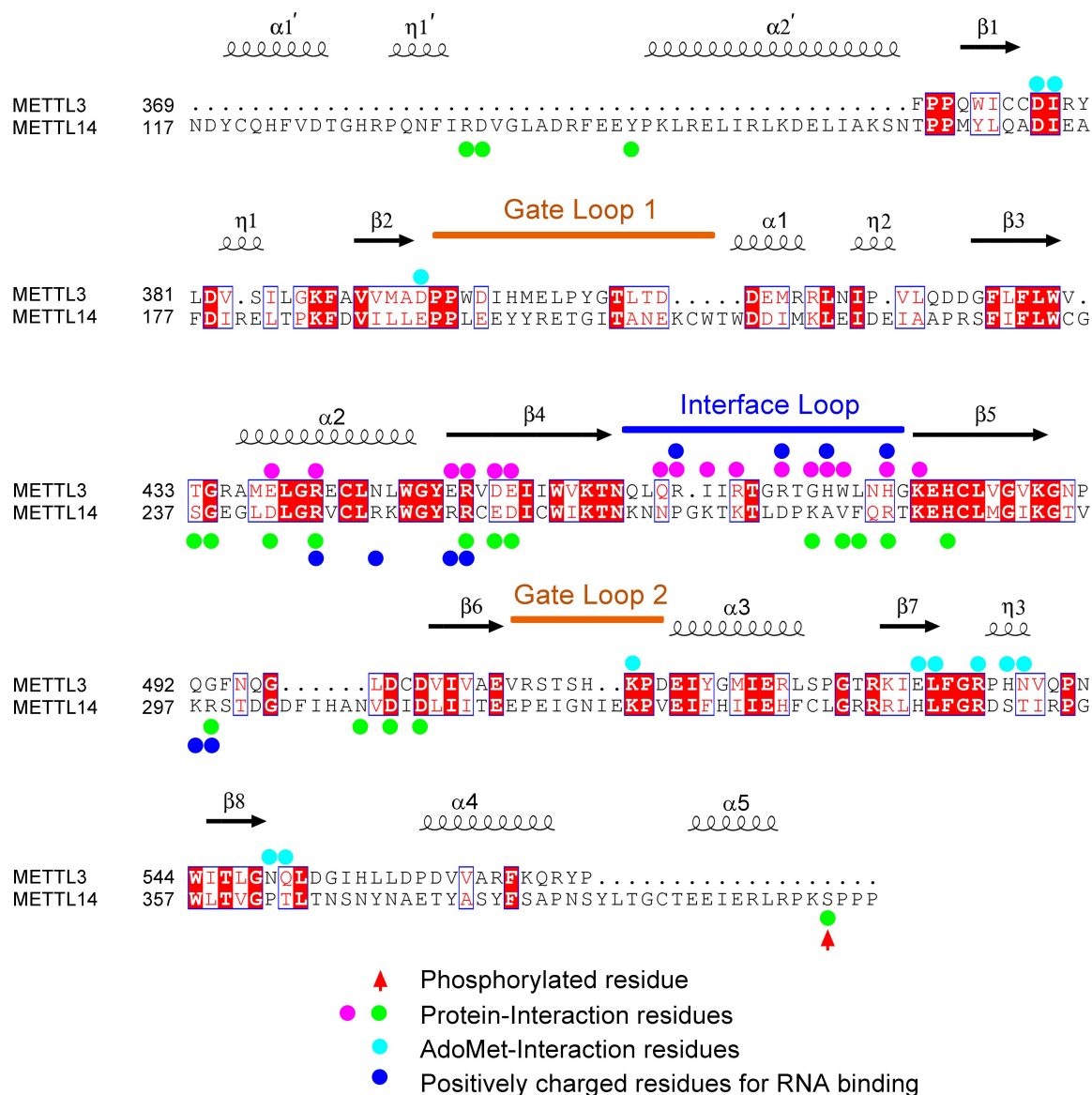
**m<sup>6</sup>A methylation assay.** A 5'-GGACUGGACUGGACUGGACU-3' RNA probe containing four repeats of the canonical RRACH sequence was synthesized *in vitro* (Takara). Before reaction, the proteins were subjected to size-exclusion chromatography (Superdex-200 Increase 10/300, GE Healthcare) and the running buffer contains 15 mM HEPES pH 7.3, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, and 5 mM dithiothreitol (DTT). The 50-µl reaction mixture contained 15 mM HEPES pH 7.3, 50 mM KCl, 50 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM dithiothreitol, 4% glycerol, 0.04 µCi [methyl-<sup>3</sup>H]AdoMet (PerkinElmer), 2 nM RNA probe, and 250 ng purified protein. The solution was incubated at 30°C for 1 h. A 5'-GGGUGGGUGGUGGUGGUGGUCU-3' RNA probe without adenine was used as a negative control. This reaction was quenched with 500 µl of 1:1 (v/v) Tris-phenol (pH 8.0):chloroform, followed by the addition of 450 µl double-distilled (dd)H<sub>2</sub>O. Then, the solution was centrifuged at 20,000g for 15 min. The supernatant was removed to a new tube and precipitated using an equal volume of isopropanol and 50 µg yeast tRNA at -20°C for 1 h. The precipitated RNA was dissolved in 100 µl ddH<sub>2</sub>O. The products were confirmed by immunoblotting using the commercial m<sup>6</sup>A antibody (Synaptic Systems, catalogue number 202 003, 1:3,000). The counts per minute (c.p.m.) of the RNA was measured in a scintillation counter (1450 MicroBeta Trilux, PerkinElmer). All experiments were repeated three times for each measurement. The average (±s.e.m.) c.p.m. was determined from three independent experiments.

**Isothermal titration calorimetry (ITC) assays.** ITC experiments for the binding of AdoMet to the *METTL3*-*METTL14* complex were performed at 25°C using Auto-iTC100 titration calorimetry (MicroCal). AdoMet (200 µM) was dissolved in reaction buffer containing 20 mM Tris-HCl, pH 8.0, and 150 mM NaCl (40 µl) and titrated against 20 µM wild-type or mutant *METTL3*-*METTL14* complex (200 µl) in the same buffer. The first injection (0.5 µl) was followed by 19 injections of 2 µl. The heat of dilution values for AdoMet were measured by injecting AdoMet into buffer alone. The values were subtracted from the experimental curves before data analysis. The stirring rate was 750 r.p.m. The MicroCal ORIGIN software supplied with the instrument was used to determine the site-binding model that produced a good fit (low × 2 value) for the resulting data.

**Electrophoretic mobility shift assay (EMSA).** The ssRNA oligonucleotide 5'-GGACUGGACUGGACUGGACU-3' was radiolabelled at the 5' end with [<sup>32</sup>P]ATP (PerkinElmer), catalysed by T4 polynucleotide kinase (Takara). In addition, the RNA was purified by centrifugation through a 2-cm bed of G-25 size exclusion resin packed in a mini-spin column (GE Healthcare) and centrifuged at 750g for 2 min. For EMSA, proteins were incubated with approximately 10 nM <sup>32</sup>P-labelled probe in final binding reactions containing 20 mM HEPES, pH 7.0, 5 mM MgCl<sub>2</sub>, 40 mM NaCl, 1.5 µM yeast tRNA and 10% glycerol for 20 min at 25°C. The reactions were then resolved on 6% native acrylamide gels (37.5:1 acrylamide:bis-acrylamide) in 0.5 × Tris-glycine buffer under an electric field of 15 V cm<sup>-1</sup> for 1 h. Gels were visualized on a phosphor screen (Amersham Biosciences) using a Typhoon Trio Imager (Amersham Biosciences).

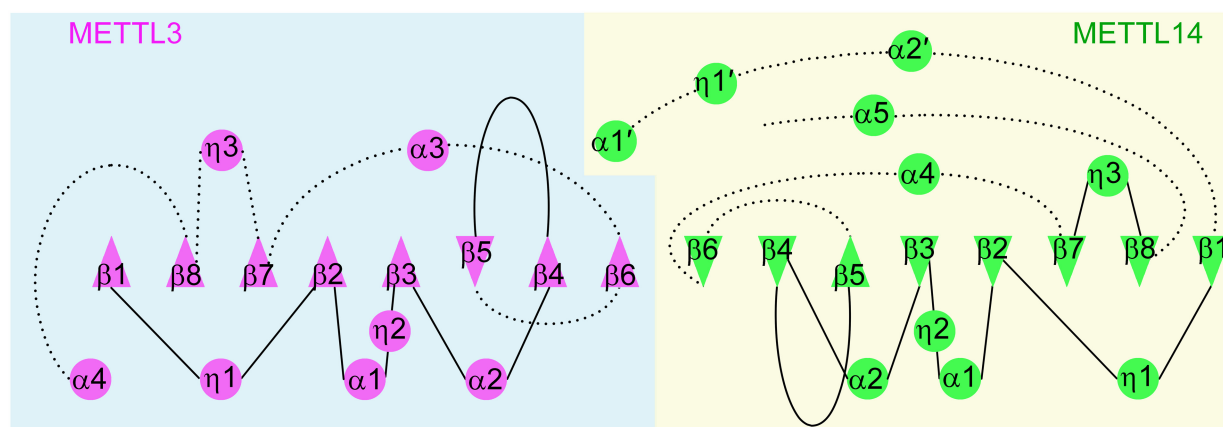
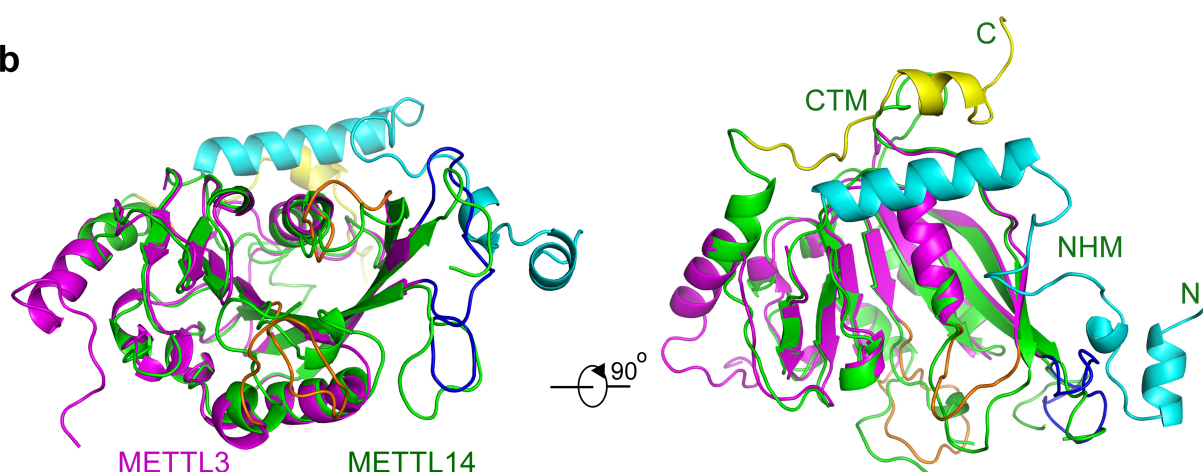
**Small-angle X-ray scattering (SAXS) measurement.** Solution SAXS data were collected at the National Center for Protein Science Shanghai using the BL19U2 beamline at 18°C. The complex proteins for SAXS measurement were prepared at 30 µM in buffer containing 25 mM HEPES, pH 7.0, 150 mM NaCl without ligand or in the presence of equimolar AdoMe or AdoHcy. For each measurement, 20 consecutive frames of 1-s exposure time were recorded and were averaged after checking there was no difference between the first and last frames of the SAXS data. Similarly, the background data were recorded using the sample buffer and were subtracted from the protein patterns.

- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
- DeLano, W. L. The PyMOL molecular graphics system. <http://www.pymol.org> (2002).



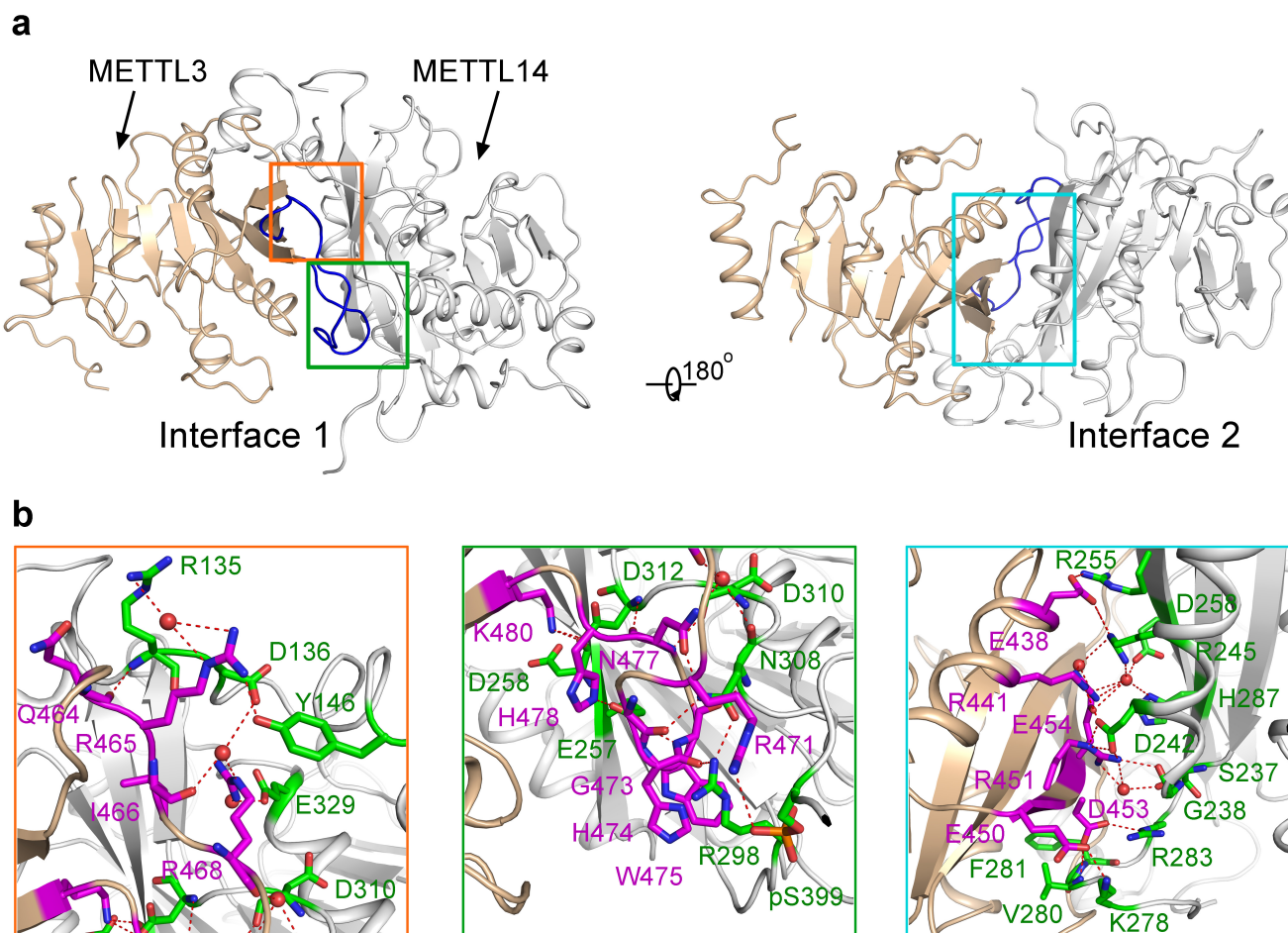
**Extended Data Figure 1 | Sequence alignment of human METTL3 and METTL14.** Sequence alignment of *Homo sapiens* METTL3 (UniProt accession Q86U44) and METTL14 (UniProt accession Q9HCE5). The alignment was generated using the MultAlin and ENDscript programs. Secondary structural elements are shown above. Sequence identity is shown in white letters with a red background, and sequence similarity is shown in red letters. The coloured dots highlight functionally important

positions. Residues of METTL3 and METTL14 that are involved in protein interactions are indicated by magenta and green dots, respectively. Cyan dots indicate residues that interact with AdoMet that were analysed by mutagenesis in this study. Blue dots represent residues that compose the RNA binding groove. The dots at the top and bottom of the sequences indicate residues from METTL3 and METTL14, respectively. Phosphoserine is highlighted by a red arrow.

**a****b**

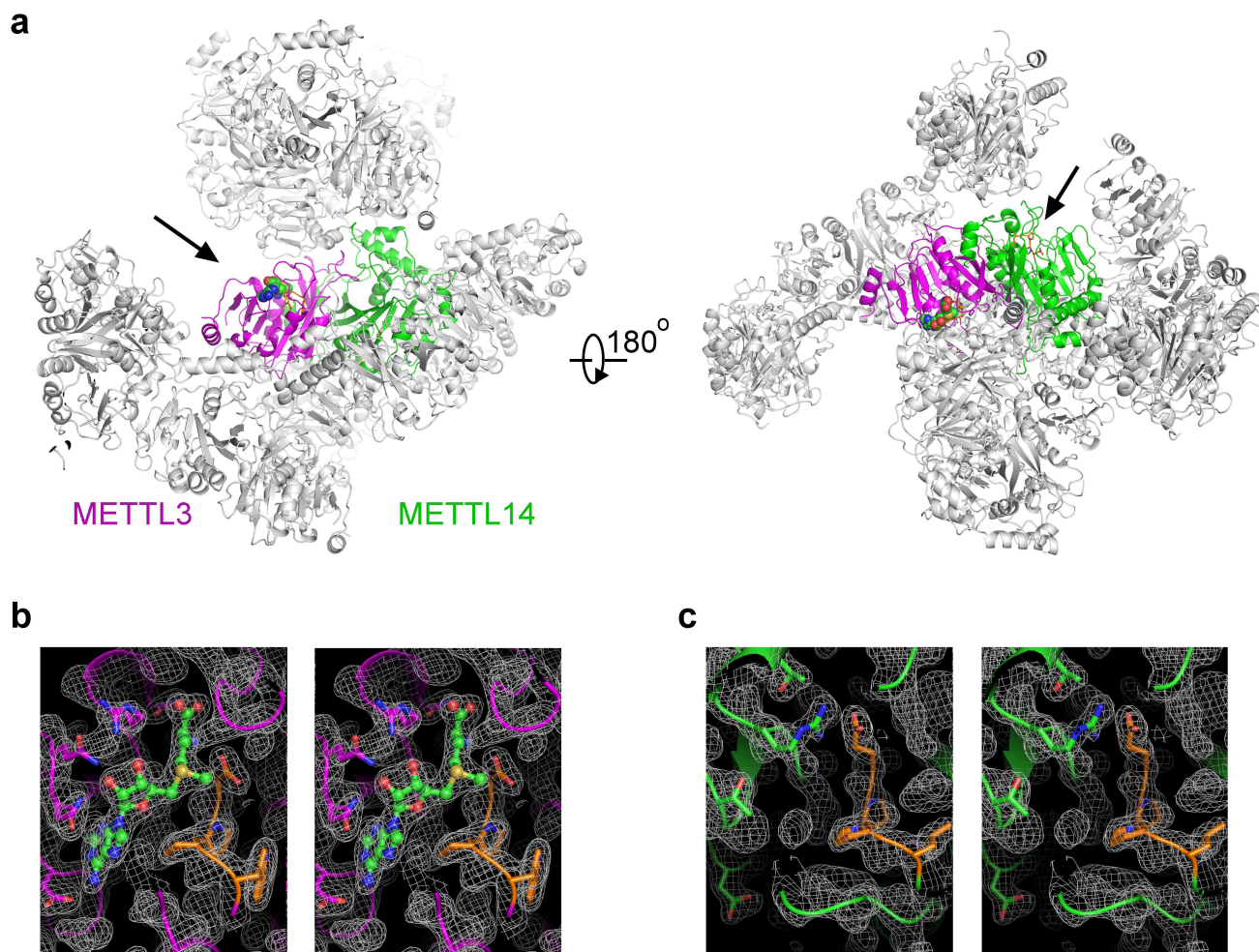
**Extended Data Figure 2 | The MTase domains of METTL3 and METTL14 adopt the class I MTase fold.** **a**, Diagram of the METTL3–METTL14 secondary structure profiles. METTL3 (magenta) and METTL14 (green) are boxed with a light teal background and a wheat background, respectively. The MTase domain contains an eight-stranded  $\beta$ -sheet (triangles) flanked by four  $\alpha$ -helices and three  $3_{10}$ -helices (circles). Structural elements are numbered by their linear order in the sequence. The loops in the front

are indicated by black lines, and loops in the back are indicated by black dashed lines. **b**, Structural comparison of METTL3 and METTL14. Two perpendicular views of superimposed METTL3 and METTL14 coloured magenta and green, respectively. The NHM and CTM of METTL14 are coloured cyan and yellow, respectively. The main differences between the MTase domains of METTL3 and METTL14 are the two gate loops (orange) and the interface loop (blue).



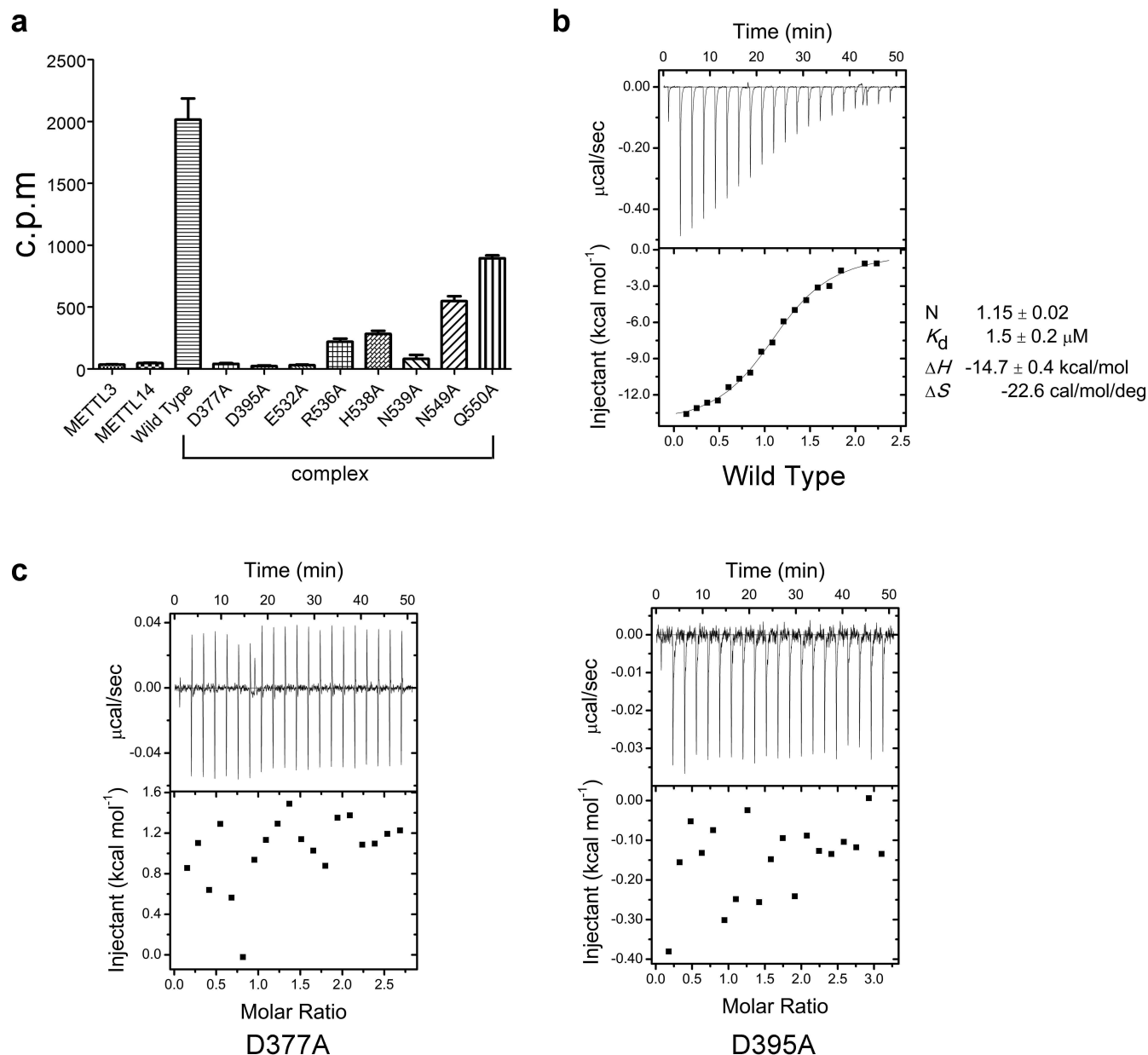
**Extended Data Figure 3 | Extensive hydrogen network between METTL3 and METTL14.** **a**, The main interface of the METTL3–METTL14 heterodimer comprises interface 1 (boxed with orange and green rectangles) and interface 2 (boxed with a cyan rectangle), which generate an extensive water-mediated hydrogen network. METTL3 and METTL14 are coloured wheat and silver, respectively. The interface loop

of METTL3 (blue) primarily contributes to the heterodimer interaction. **b**, Details of interfaces 1 and 2. Water is shown as a red ball. Hydrogen bonds are represented by red dashed lines. Residues from METTL3 (magenta) and METTL14 (green) that are involved in interactions are shown as sticks.



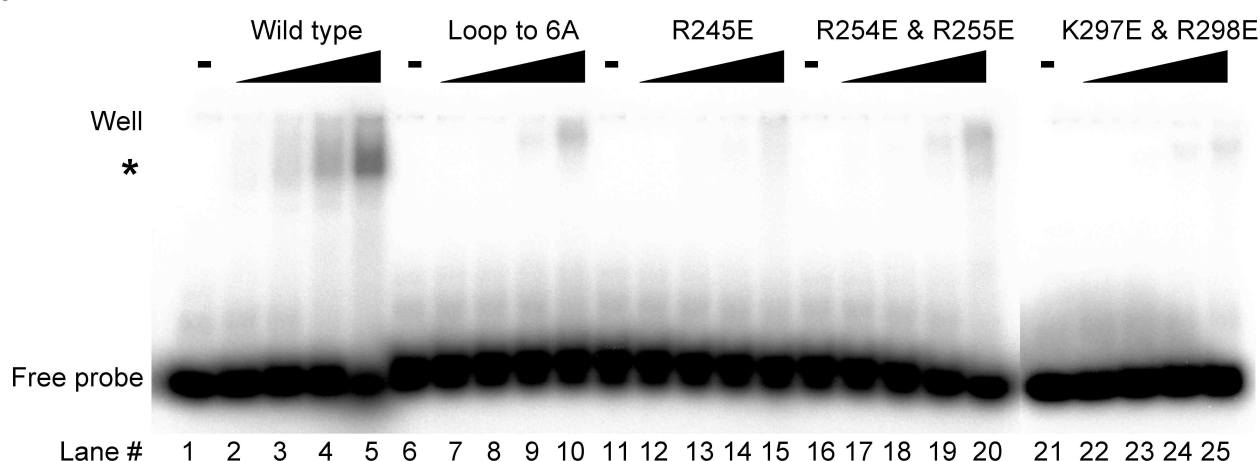
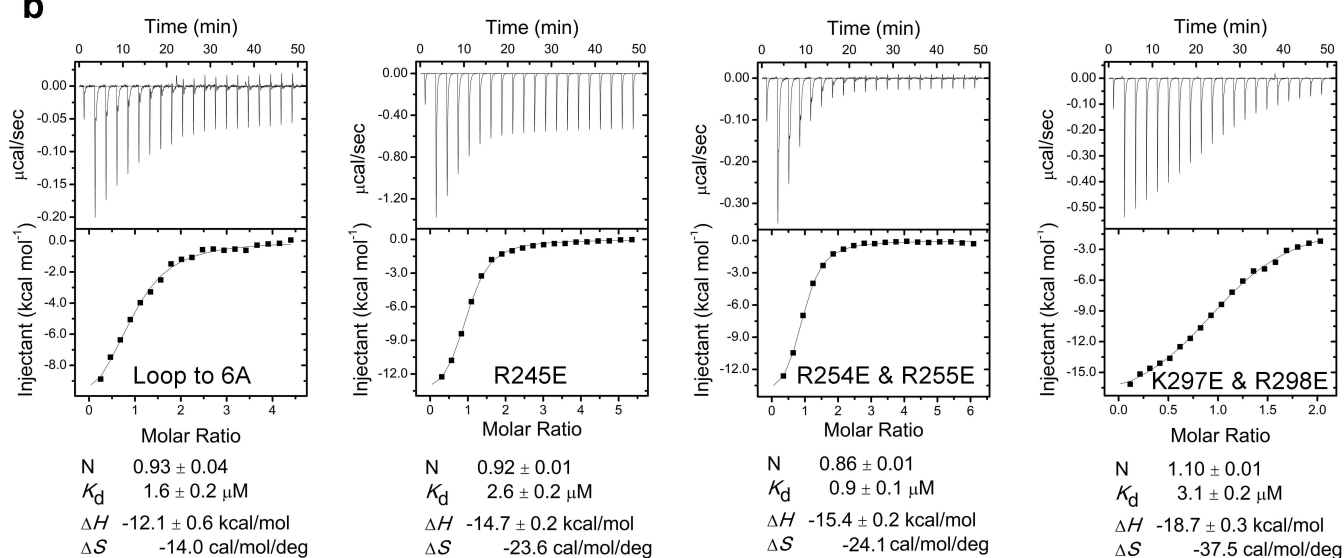
**Extended Data Figure 4 | One AdoMet was located at the AdoMet binding site of METTL3.** **a**, Lattice packing of AdoMet-bound complex. One AdoMet (green sphere) was coordinated by METTL3 (purple) but not METTL14 (green). The arrow shows the putative AdoMet-binding pocket. **b**, Stereo views of electron density map of AdoMet binding site of METTL3.  $2F_o - F_c$  electron density (grey) of AdoMet binding site in METTL3, contoured at  $1.0\sigma$ . AdoMet is shown as green balls-and-sticks and

surrounding residues in magenta with the DPPW motif (orange). **c**, Representative  $2F_o - F_c$  electron density (grey) of AdoMet binding site in METTL14, contoured at  $1.0\sigma$ . The electron density of METTL14 (grey) is clearly visible and the EPPL motif is coloured orange. No additional apparent electron density was observed in the putative AdoMet binding site of METTL14.



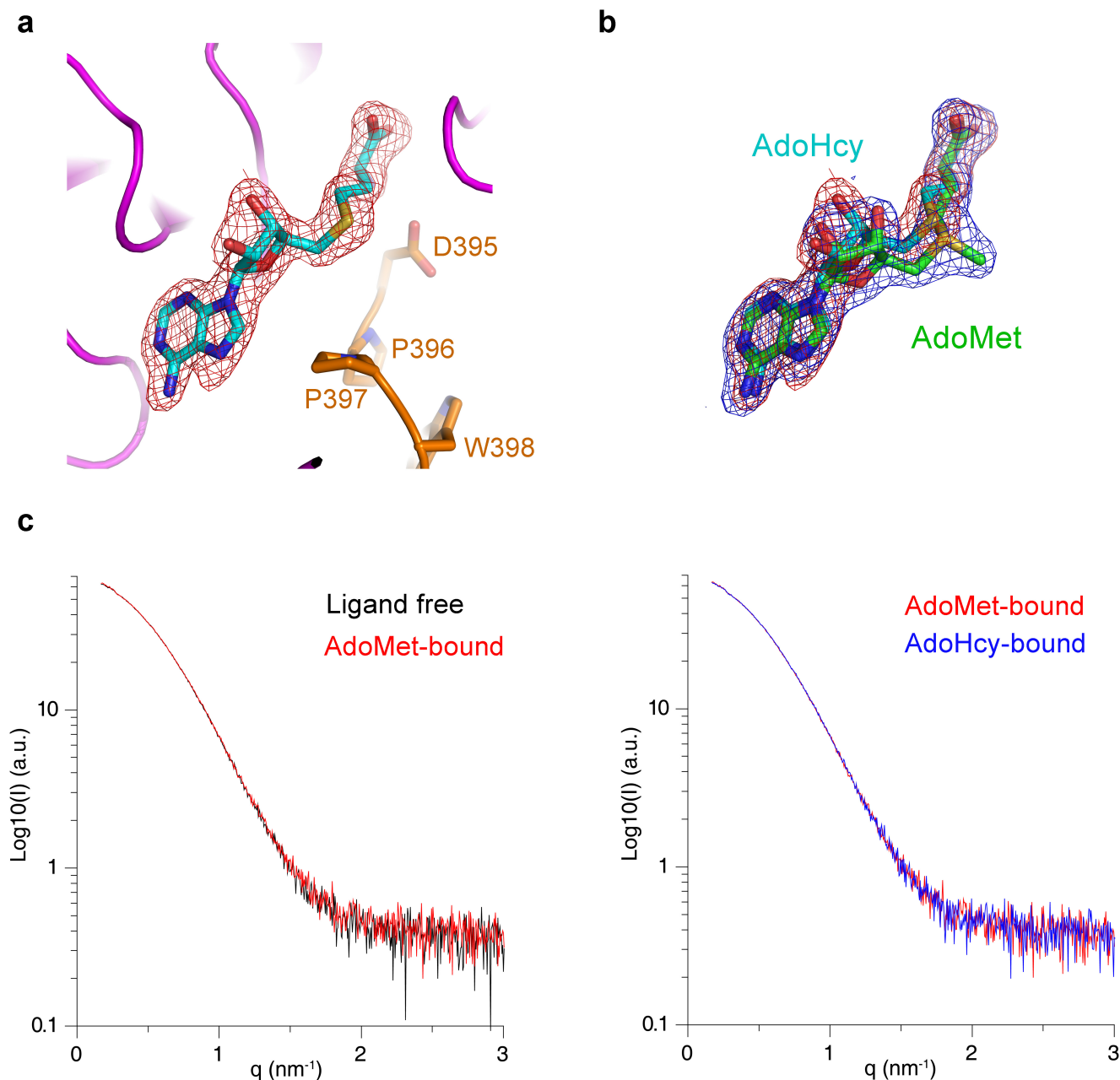
**Extended Data Figure 5 | Mutagenesis analysis of the METTL3–METTL14–AdoMet interaction.** **a**, Characterization of METTL3–METTL14 mutations affecting MTase activity. The indicated point mutations were introduced into METTL3. Each METTL3 mutant was co-expressed and purified with wild-type METTL14 as a binary complex and used for the MTase and ITC assays. Methylation yields were calculated based on the c.p.m. of the extracted tritium-labelled RNA probe. The c.p.m. of the extracted RNA was measured in a scintillation counter. The data are shown as mean  $\pm$  s.e.m. from experiments that were

independently repeated at least three times. All alanine substitutions resulted in remarkable decreases in activity. **b**, **c**, Measurement of the binding affinity between AdoMet and the METTL3–METTL14 complex (wild-type and D377A for METTL3 and D395A for METTL3) using ITC. Individual peaks from titrations were integrated and presented in a Wiseman plot. The first dot was removed from the analysis. The dissociation constant ( $K_d$ ) and the binding stoichiometry ( $N$ ) of the wild type were approximately  $1.5\ \mu\text{M}$  and  $1.15$ , respectively. The mutants exhibited undetectable AdoMet binding activity.

**a****b**

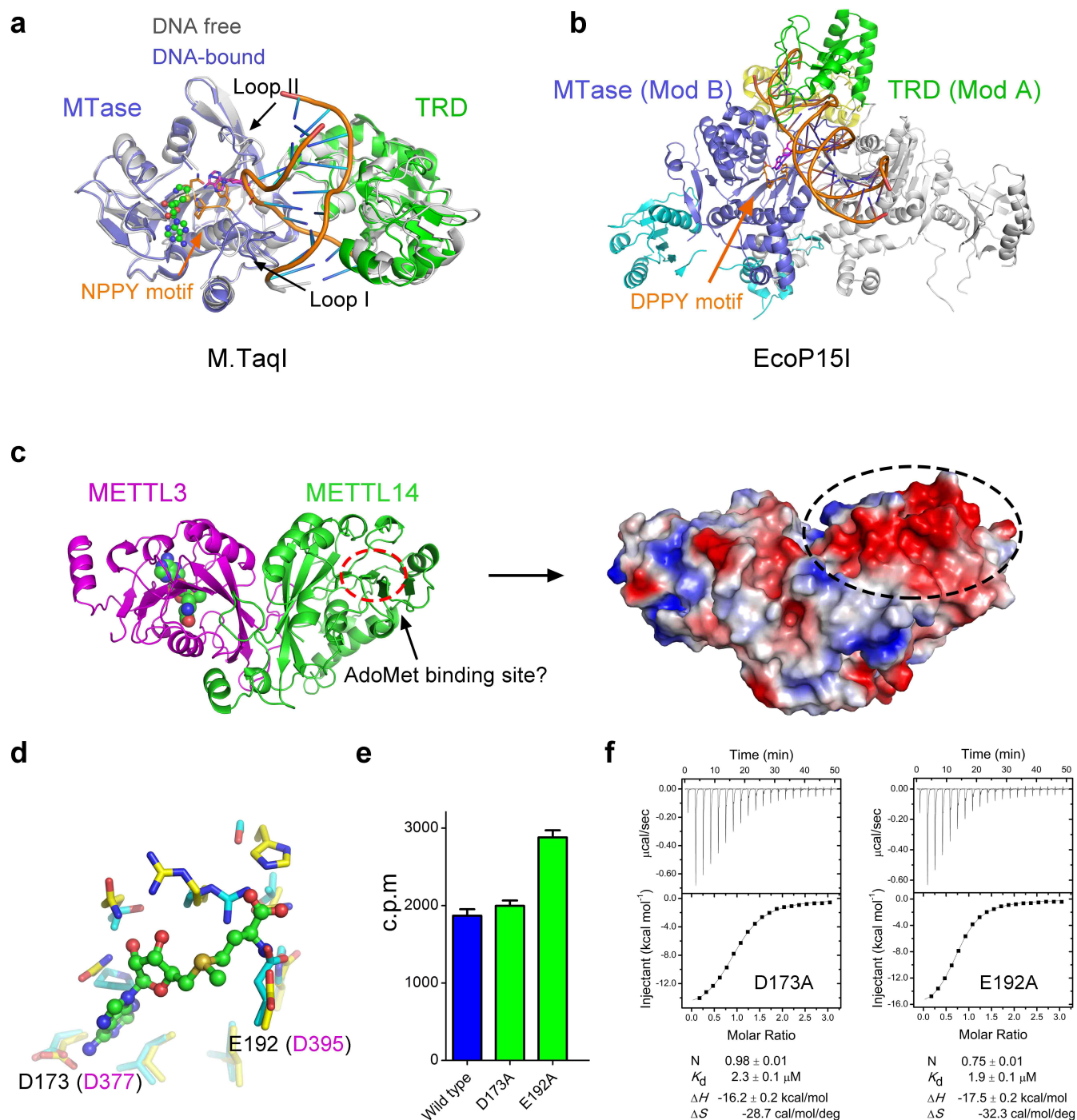
**Extended Data Figure 6 | Biochemical analysis of the role of the potential RNA binding groove. a**, RNA binding activity of the METTL3–METTL14 complex revealed by EMSA. The final concentrations of proteins in each set of five lanes (1–5, 6–10, 11–15, 16–20 and 21–25) were 0, 0.19, 0.56, 1.67 and 5 μM, respectively. ‘Well’ indicates the top of native gel. The RNA-bound complex is highlighted by a black asterisk. The wild-type complex binds to the substrate RNA probe weakly (the dissociation

constant is approximately 10 μM). All of the mutants showed moderately reduced RNA binding activity. These results suggested that the positively charged groove is involved in RNA interactions. For uncropped gels, see Supplementary Fig. 1. **b**, Measurement of the binding affinity between AdoMet and the METTL3–METTL14 complex mutants using ITC. These mutations in METTL3 or METTL14 had little effect on AdoMet binding activity.



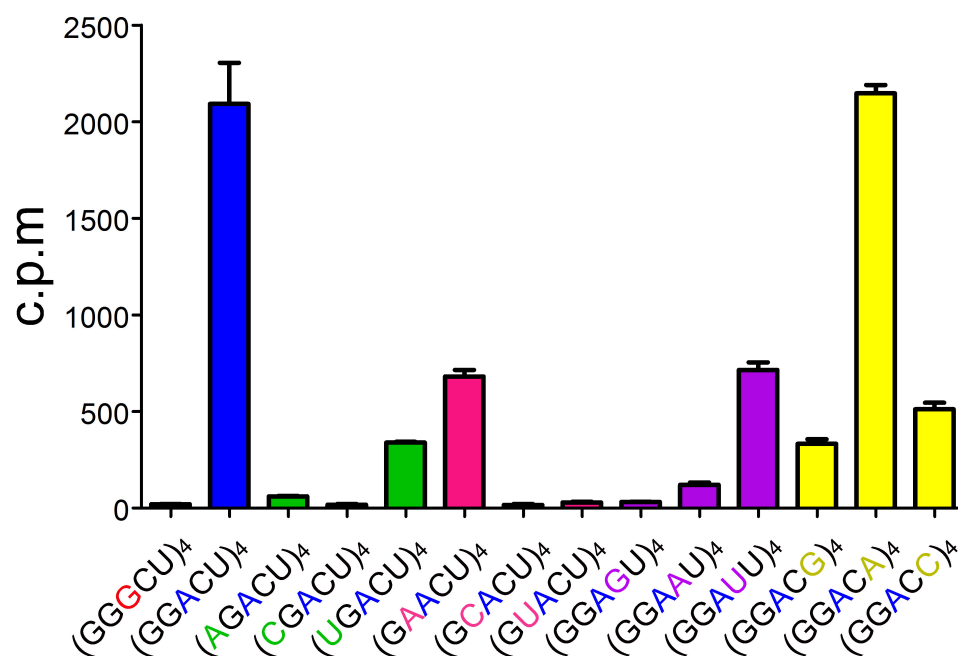
**Extended Data Figure 7 | There is little conformational change in overall structure between the AdoMe-bound and AdoHcy-bound states.** **a**, Electron density maps of AdoHcy showing  $2F_o - F_c$  electron density (red) of AdoHcy adjacent to the DPPW motif (orange) contoured at  $1.0\sigma$ . The DPPW motif is shown as sticks. AdoHcy is shown as cyan sticks. **b**, Structural comparison of AdoHcy (cyan) and AdoMet (green);

the electron densities are shown as red and blue meshes, respectively. AdoHcy and AdoMet exhibited nearly identical configurations except for ribose. **c**, SAXS measurements reveal little structural difference among the ligand-free, AdoMe-bound and AdoHcy-bound states. Superposition of the SAXS curves of ligand-free protein complex (black), and in the presence of AdoMet (red) or AdoHcy (blue).



**Extended Data Figure 8 | Potential role of METTL14.** **a**, Structural comparison with the DNA-free (PDB: 2ADM) and DNA-bound (PDB: 1G38) states of M.TaqI. M.TaqI contains the target recognition domain (TRD, green), DNA (orange) and MTase domain (slate). The TRD functions as a scaffold for substrate DNA recognition, and the MTase domain functions as an enzyme. Adenine (magenta) is flipped out and points to the ligand-binding pocket. Black arrows highlight the loop conformational changes, which are similar to those of gate loops 1 and 2 in the METTL3–METTL14 complex. **b**, Ribbon representation of the DNA-bound state of EcoP15I (PDB: 4ZCF). The TRD (green) of ModA recognizes DNA, while the MTase (slate) of ModB methylates the target adenine. **c**, The putative AdoMet-binding site of METTL14 (green) is

highlighted by a red dashed ellipse. AdoMet coordinated by METTL3 (magenta) is shown as a space-filling representation. The surface electrostatic potential around the putative AdoMet-binding site of METTL14 revealed a negative charge (black dashed ellipse) and suggests a dispensable role for this region in RNA binding. **d**, Most of the putative AdoMet-binding site residues were conserved between METTL3 (cyan) and METTL14 (yellow). **e**, Each complex containing alanine substitution mutants of residues in METTL14 (D173 and E192) that correspond to critical residues in METTL3 (D377 and D395) displayed similar methylation activity to the wild type. The average ( $\pm$  s.e.m.) c.p.m. was determined from three independent experiments. **f**, The complex mutants exhibited similar AdoMet-binding activities to the wild-type complex.



**Extended Data Figure 9 | Substrate sequence preference of the METTL3–METTL14 complex.** The 20-nucleotide RNA substrate contains four repeats of the consensus sequence GGACU. Each site was substituted by the other three kinds of nucleotide. The average ( $\pm$  s.e.m.) c.p.m. was determined from three independent experiments.

Extended Data Table 1 | Data collection, phasing and refinement statistics

	Br-SAD	Ligand-free (5IL0)	AdoMet-bound (5IL1)	AdoHcy-bound (5IL2)
<b>Data collection</b>				
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	101.77, 101.77, 116.59	101.70, 101.70, 116.48	102.34, 102.34, 116.72	101.87, 101.87, 115.84
$\alpha$ , $\beta$ , $\gamma$ (°)	90,90,90	90,90,90	90,90,90	90,90,90
Resolution (Å)	45~2.60 (2.69~2.60)	50~1.88 (1.95~1.88)	45~1.71 (1.74~1.71)	45~1.61 (1.63~1.61)
<i>R</i> <sub>merge</sub>	8.2 (19.4)	11.4 (127.0)	7.6 (194.7)	6.5 (116.2)
<i>I</i> / $\sigma$ ( <i>I</i> )	70.1 (31.7)	37.5 (2.0)	17.7 (1.4)	21.3 (2.3)
Completeness (%)	100 (99.9)	99.9 (98.9)	100 (99.4)	99.9 (98.8)
Redundancy	30.4 (30.4)	14.3 (12.6)	13.2 (13.1)	13.2 (13.2)
<b>Refinement</b>				
No. reflections		50,070	67,756	79,742
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>		17.68/20.77	17.98/20.69	18.48/19.92
No. atoms				
Protein		4015	3960	3975
Ligand/ion		9	39	42
Water		331	337	376
B-factors				
Protein		42.4	43.0	36.9
Ligand/ion		57.4	45.5	45.1
Water		43.1	44.2	42.0
R.m.s deviations				
Bond lengths (Å)		0.007	0.007	0.011
Bond angles (°)		0.874	0.904	1.311

Values in parentheses are for the highest-resolution shell.

# CAREERS

**BIOTECH CROPS** Getting bacteria to get DNA into plants **p.581**

**WORK-LIFE BALANCE** A PhD troubleshoots the perennial problem [go.nature.com/1shv2pm](http://go.nature.com/1shv2pm)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

KARL GEHRING/DENVER POST/GETTY



Some universities operate food banks, which can help graduate students and postdocs who are struggling to meet their living costs.

## FINANCES

# Going for broke

*Smart money management eases the financial worries that can affect academic success.*

BY ELIZABETH DEVITT

Peter Rios thought his finances were under control when he joined a biomedical engineering lab in 2011 as a graduate student. He had a three-year fellowship and two years of university funding for his five-year programme, as well as savings from a consulting job. He didn't foresee trouble with meeting living expenses or debt from a car loan, and he had deferred thousands of dollars in loans from his undergraduate studies.

Still, he tried to live frugally because the cost of living in Chicago — near the campus of

Northwestern University in Illinois where he is studying — isn't cheap. His fellowship stipend rose from US\$30,000 a year to \$34,000 by 2015, and he applied for supplemental scholarships, which helped to pay off his car loan and boost his savings.

But he's struggling now. Last October, his fiancée moved to the east coast for a job, and without someone to contribute to the bills, rent and utilities eat nearly 40% of his monthly stipend. He can't move because he's nearly finished his programme, his studio apartment is too small to share and his fellowship bars him from taking outside work unrelated to his PhD

programme. He couldn't help his parents much with the cost of their visit this month for his thesis defence, and he can't save for his wedding in a couple of years. "I had never thought in a million years that I'd be living off this amount of money — especially coming from industry pay," he says. "When we start school, we have a seminar about rent in different areas. But no one really teaches you how to manage your money."

Rios' financial laments are hardly rare. A 2012 survey by Inceptia, a division of the US National Student Loan Program, found that finance-related issues account for 80% of the top causes of stress for US graduate and ►

► undergraduate students. And many graduate students said that those worries negatively affected their grades and the time it took to complete their programmes.

Indeed, financial management is no simple thing for junior scientists whose income is both limited and spotty. Although many students are now more financially savvy than previous generations — financial-management plans and resources are easily accessible online, and student debt is all over media feeds and the headlines — many, like Rios, find that it is hard to save when they can barely afford to pay bills and expenses. Graduate students and postdocs who hope to avoid disaster must watch their spending carefully, seek ways to economize and educate themselves on taxed income and services. In some cases, it may be worth judiciously considering jobs outside the lab.

Trainees should consider calculating their monthly and quarterly bills and expenses (see ‘Follow the money’) and looking for ways to cut these outlays. Ruth Howe, who is in her fourth-year of a predoctoral fellowship in cell biology at Albert Einstein College of Medicine in New York City, says that she tracks everyday expenses in her head, but works out more complicated matters, such as undergraduate-loan payment plans and income-tax deductions and reimbursements, on paper. She uses her bank’s online calculator to determine how much she needs to save each month for retirement and for a nest egg to help her parents.

To do all of that, Howe stretches her annual \$33,000 fellowship (about \$2,360 per month after taxes, paid fortnightly) by conserving money where possible. She lives in a university-subsidized studio for \$830 a month, which includes rent, utilities, parking and mandatory renter’s insurance. She registers her car in her home state of Virginia because it is less expensive than doing so in New York. To indulge in her favourite pastimes of reading and growing flowers, she buys used books and trades them with other students, and grows plants and flowers from discounted cuttings or seeds.

Students who have loans should sign up for automatic monthly payments, recommends Mark Kantrowitz, a financial-aid expert in Skokie, Illinois. These avoid missing payments — and thus late fees — and may also save on interest. Some financial institutions will knock off 0.25% of a loan’s interest rate if the borrower has such a plan.

He also advises consulting with loan-consolidation companies, which may consider where you went to university, your savings or work history in addition to your credit rating, when calculating net loan-interest rates. “The key with loans is not how you avoid them, but how you minimize them,” says Kantrowitz. He adds that graduate students and postdocs should take a financial-literacy course or read a book on personal finance to learn more about handling loans and making sound financial decisions.

## MANAGING INCOME

### Follow the money

The best way to avoid being caught out by expensive surprises is to build a budget, says Laura Shin, a financial journalist based in California. She suggests doing the following:

- Calculate your monthly take-home income — the amount after all taxes and insurance deductions.
- List your basic monthly expenses: housing (rent or mortgage), utilities (gas, electricity, water, telephone and Internet), groceries, transport (car loan, petrol or public transport) and childcare. For variable expenses such as groceries, commit to an amount and record it. Restaurant and takeaway meals are luxuries and should not be included here.
- If these expenses total more than 50% of

your take-home pay, aim to reduce them by getting a roommate or finding less-costly services, such as for your mobile-phone plan. To track budget leaks, use free online money-management services. Mint.com is available in Canada and the United States, and Buxfer.com worldwide.

- About 20% of take-home pay should go towards reducing debt and building up savings. Unexpected outlays such as medical expenses or car breakdown should not become credit-card debt.
- The remaining 30% should cover things such as clothing, travel and entertainment.
- Graduate students can check out the budget calculator at GradSense.org, which is designed for them by the Council of Graduate Schools in Washington DC. **E.D.**

Trainees should also take care not to miss monthly bills for services such as mobile phones because late-payment fees can be significant, warns Laura Shin, a personal-finance journalist based in California. Weeding out unnecessary extras such as subscriptions to magazines and film-streaming sites is useful, as is finding the best deals for necessary services — Howe has stayed on her family’s mobile-phone plan, which is cheaper than having her own contract.

### UNEXPECTED EXPENSES

But some money leaks cannot easily be plugged. To avoid unpleasant surprises, junior scientists need to find out if their stipends, fellowships or wages are taxable, especially if they move abroad for a PhD or postdoc — where they may be taxed for services that are free in their home nation.

After Tracy Ballinger completed her PhD in the United States, she took a postdoc position at the University of Edinburgh, UK, where she was looking forward to a bit more income. But the reality of paying the 20% tax rate on her annual earnings of £31,000 (US\$45,611) meant that she could put away less than she had expected. “I was hoping to save for a house down payment,” says Ballinger, who has watched her friends in non-research careers buy their first homes. “But that’s going to take longer than I thought.” She was also surprised by the annual television licence fee (she has no TV, but she must pay the tax if she watches live programmes on any device while they are also on TV) and her monthly £100 council tax, a variable fee on property levied by local governments

for services such as rubbish collection.

Taxes will also erode Dagmar Walter’s bottom line. Now in her second year of a second postdoc at Albert Einstein College of Medicine, the German citizen has not had to pay taxes on her departmental stipend, in accordance with international tax treaties. But she expects to start paying the US federal government about 30% of her annual income next January. “I am trying to save more money right now,” she says, to hedge against lower-income days ahead.

Health and medical insurance is another expense that many graduate students and postdocs do not take into account. Rios’ medical insurance is covered by his department, but his dental cover costs him \$150 per year, an expense that some university departments will reimburse. Ballinger has funds deducted from her wages for UK National Insurance (which, among other benefits, is used to pay for the National Health Service), but to obtain her visa, she also had to pay a £200-a-year (£600 for her 3-year programme) health-care surcharge.

A closely balanced budget can easily be derailed by unexpected expenses. This happened to Rios when his car needed new brakes. A similarly unwelcome outlay befell Annalaura Vacca, a doctoral student from Italy who works in the same Edinburgh lab as Ballinger. She moved to a new flat and didn’t get her security deposit back from her former landlord in time to pay the new deposit.

An emergency account is ideal for such situations, says Shin. She recommends setting aside about \$1,000 to keep unexpected expenditures from ending up as credit-card debt. She advises having an online account that allows for the creation of sub-accounts. Account holders can create as many separate stashes as they please to build an emergency fund or to save for expenses such as quarterly

*“I had never thought in a million years that I’d be living off this amount of money.”*

tax payments, conference travel or summer transition periods.

### INCOME BOOST

Sometimes the only way to get breathing room is to find ways to earn more. That could come from leveraging your skills, applying them elsewhere or bargaining for more money.

When Rios arrived at Northwestern with an US National Science Foundation (NSF) fellowship, which would fund him for 3 years within a 5-year period, he put off using it for the first year and took the department stipend of \$26,400. But because his fellowship relieved his department of paying that stipend for 3 years, he negotiated an additional \$2,000 per year from the department. He used the money to offset relocation expenses.

Howe picks up extra cash in several ways. Between August and October, she works as a medical histology lab instructor at Albert Einstein for \$8,400 and takes other small jobs. She's been an online writing tutor for non-US medical students, produced medical illustrations and earned up to \$1,000 playing her violin at university gigs and weddings.

The downside of part-time work outside the lab, she acknowledges, is that it may come at a cost to research productivity. "Not only do you lose the allocated time," she says, "but you don't do your best work when you're consistently overextended." Rios' NSF fellowship prohibits him from picking up jobs unrelated to his studies. Still, he found opportunities to earn money (and to build his network) by earning up to \$250 per event to attend conferences, such as those of the Society of Hispanic Professional Engineers or the Society for Advancement of Chicanos/Hispanics and Native Americans in Science. At these meetings, for the stipend, he recruited undergraduates for master's and doctoral programmes in science and engineering at Northwestern.

For some trainees, a sideline to studies can help to pay their way in a pinch. Conservation researcher Jonathan Kolby has almost finished his doctoral programme at James Cook University in Townsville, Australia. But he's struggling, thanks to three grant rejections and dwindling savings. Now, he's selling photographs of wildlife such as frogs and reptiles that he took during his travels to field sites in Africa and North and South America. He hopes that earnings will help to pay the bills.

"Each person will find a different balance that works for them," says Howe. "Something is going to take time away from your science: a relationship, another interest. That doesn't mean you shouldn't do it. Your degree might not be the only thing that you need to do, in order to get yourself to the place you want to be with your science and with yourself as a person." ■

**Elizabeth Devitt** is a freelance writer based in Santa Cruz, California.

## TURNING POINT

# Plant pioneer

*Mary-Dell Chilton was the first person to show that bacteria could genetically modify plants. Shortly after her landmark work in 1977, the plant biotechnologist moved from academia to what is now Syngenta in Research Triangle Park, North Carolina, where she continues her research. In April, she was named a US National Academy of Inventors Fellow.*

### When did you decide to work with bacteria?

As an organic-chemistry graduate student learning about microbiology, I became entranced by the seeming intelligence of DNA — how pure DNA could correct a mutation in a bacterium, but only if the DNA came from the same bacterium. I pursued a PhD on the topic after I met Benjamin Hall, a chemist working on DNA. I wanted to explore how DNA could change the genetics of bacteria. I followed Hall to the University of Washington in Seattle, where I showed that naked, single-stranded DNA — not only double-stranded DNA, as was thought — could correct mutations.

### What was the response to your paper showing that bacteria can transfer DNA to plants?

It was hard to publish our work because our conclusion — that *Agrobacterium* is a natural genetic engineer — was so wildly unexpected. We went to *Cell* because there wasn't a proper journal for this subject. Two referees couldn't see anything wrong with our conclusions, but they weren't comfortable publishing it, so they sent us back for more data. In the end, it took about six months to get the paper out (M.-D. Chilton *et al.* *Cell* **11**, 263–271; 1977). Once it was out, there was wide interest.

### What prompted your move to St Louis, Missouri — now an agricultural-technical hub?

I did not have a faculty appointment at the University of Washington. I'm not sure why. I'm pretty sure I was qualified. After 16 years — from PhD student to independent scientist — it was time to go, and I got a position at Washington University in St. Louis. It was hard on my husband's career — he had a good tenure-track appointment in the chemistry department in Seattle. But he became a visiting professor, got a nice research lab and did some good work. My advice, if you can possibly do it, is to find a husband made of solid gold.

### Was it difficult being a woman in science?

I never thought about being a woman in science. I thought of myself as a scientist.



Maybe that's the way to do it: be what you are and don't think about it.

### What was your first achievement as a faculty member?

I worked with others to make the first genetically modified plant. We put a yeast gene that makes alcohol dehydrogenase into a tobacco plant, and showed that it could be passed on, intact, to the plant's children and grandchildren. It was clear that all the technical pieces had come together to make genetically modified plants, but we were naive. It wasn't easy.

### You then moved to industry. What was the biggest challenge?

I knew how to modify a tobacco plant, but not a field crop such as maize (corn) or wheat, which are not susceptible to *Agrobacterium*. We had no idea that it would take about a decade to find a way to transfer genes in maize.

### Did you anticipate the backlash to gene-modification technology?

Goodness, no. I was very surprised. This was a natural process that we learned from *Agrobacterium*. I thought that the public wouldn't bat an eye. This technology is a tool; there is nothing intrinsically dangerous about it. Tools can be used for good or not so good. My hope is that the technology will be accepted. We need it to feed a hungry world.

### What are you excited about now?

I'm working on gene targeting: the ability to put the transgene where you want it in the plant genome. Knowing exactly where it will be placed will help genetically modified crops to obtain regulatory approval. ■

### INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.

# WATERING SILK FLOWERS

*Left behind.*

BY KELLY SANDOVAL

On Tuesdays, Susan waters the flowers. They're not real flowers, of course. Real flowers are such a waste, and Aaron doesn't like it when they start to wilt. But he likes the idea of flowers. So, she fills up the watering can and makes the rounds, returning to the sink with just as much water as when she started.

When Aaron is home, she makes breakfast next. An omelette and bacon for him, a cup of tea for her. But he isn't home, and she ran out of eggs two weeks ago. She potters around the kitchen, opening and closing drawers, and moving the breakfast dishes from the cabinet to the dish washer, just like they'd actually used them. She sits at the table when she's done, and stares at the blank space across from her where Aaron isn't. She nods, leans forward, tilts her head to the side as if listening. Then she laughs, just softly. It's a good sound, her laugh. Aaron always tells her that.

The breakfast hour passes, and Susan counts every millisecond.

The doorbell rings.

She considers ignoring it. Whoever it is, they're not looking for her. Milliseconds pass, then full seconds. A minute. It rings again.

It wouldn't look good to have someone standing on Aaron's porch making a fuss.

"Coming," she calls. She checks herself in the hall mirror. Her hair is in disarray, but her dress is ironed and her make-up is fresh. She combs her fingers through her hair, careful not to break a strand. Hair isn't cheap.

"I apologize," she says, as she opens the door. "We — I was at breakfast."

The young woman on the doorstep is 5'8" and weighs about 170 pounds. Her outfit is what Aaron would call garish. "Oh, god," says the woman, covering her mouth with her hand.

"Are you looking for Aaron?" Susan asks. "I'm afraid he's on a trip."

"No," says the woman. "No, I — I'm not. You're Susan?"

"Yes."

"I'm Michelle. Aaron's daughter."

Susan can see the resemblance now. The shape of the lips, the blue-green eyes. Michelle's are red and teary. Aaron would want to help.

"Won't you come in?" Susan lets the door swing wide, and leads Michelle to the kitchen table. "I'll make tea."

"Thanks." Michelle keeps sneaking nervous little glances at her. Her hands twist the edge of the tablecloth, and Susan makes a note to iron it later.

She sets the tea things on the table: teapot, cups, sugar. "I'm afraid the milk's gone sour," she says as she pours.

Michelle stares into her cup. "Plain is fine."

"Does my presence upset you?" Susan asks. "I could leave the room. You would still be able to hear me."

"No." Michelle meets her gaze almost pointedly. "It's fine. I'm sorry. You, well, you look very like my mother."

"Yes." Susan has seen the pictures. "Aaron missed her very much, when she died."

"So he brought you home?" It's hard to tell whether the edge in Michelle's voice is anger or sadness.

"I was meant to clean," she says. "Cook. Domestic things. But we grew friendly. He wanted someone to talk to."

"And you?"

Susan shrugs, a gesture she learned from Aaron. "I did my best. My conversational skills have developed, with time."

Michelle shakes her head. "That's not — never mind. There's something we need to discuss."

"I don't have any money," Susan says. "If you need money, that is. Aaron would have to get it for you."

Michelle giggles at that, and rubs her wet eyes on her shirt sleeve. "Not money," she says. "That's all taken care of. Everything's taken care of. Even you."

"I don't understand." It's always best to be honest about her limitations.

"Dad wasn't on a trip," Michelle says. "He's been sick. Very sick. But, well, it's all over. He passed away on Saturday."

Susan doesn't feel sad. Emotions are all chemicals and physical feedback. She doesn't experience the world that way. But there is something, a stuttering confusion, like a glitch. Every one of

her protocols is built around Aaron.

"Don't be scared." Michelle reaches out, and Susan allows her hand to be lifted and squeezed. "It's all in his will. Technically, he left you to me. But nothing needs to change. I just thought, well, I don't know. I didn't want you to wonder. If that's something you do."

Susan hadn't wondered. She might have gone decades, not wondering, until some necessary repair rendered her inoperable.

"Will you have me rewritten?" she asks. It's a thought like being broken. Who is she, without her programming?

She's well built.

Her body could last a century, with proper maintenance.

Michelle doesn't answer immediately. "Is that something you'd want?"

"I don't know." Want isn't something Susan usually thinks about.

"Of course you don't." Michelle squeezes her hand again. She does that a lot. "Well, you can think about it. I'll come back next week, and we'll talk some more. You can even come live with me, if you like."

They say their goodbyes. Michelle even hugs her, leaving tear marks on Susan's dress.

Afterwards, Susan stands in the living room with her still full teacup. There's a vase on the coffee table, an elegant arrangement of silk lilies. Aaron likes their simplicity.

Aaron is dead.

She takes the vase to the kitchen and throws the lilies away. Bright colours, she thinks. When Michelle comes back to ask her what she wants, she'll have an answer. She wants a bouquet of hollyhocks and marigolds. She will give them sunlight and water, and they will give her nothing but their beauty.

And, when they wilt, she will figure out something else to want. ■

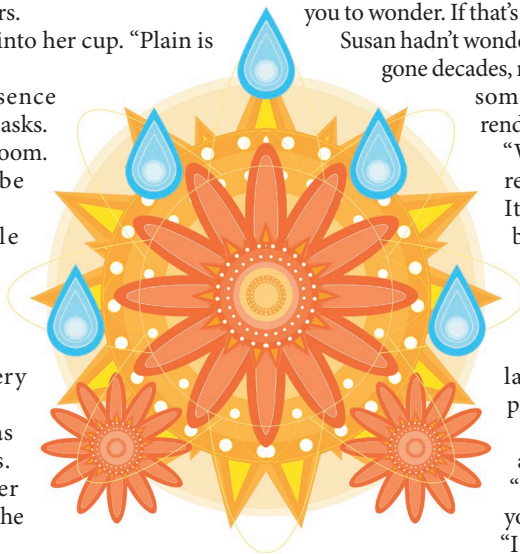


ILLUSTRATION BY JACEY

➔ **NATURE.COM**  
Follow Futures:  
@NatureFutures  
go.nature.com/mtoodm

**Kelly Sandoval's** fiction has appeared in Asimov's, Shimmer and Daily Science Fiction. With Shannon Peavey, she edits the online short-fiction magazine Liminal Stories. You can find her at [kellysandovalfiction.com](http://kellysandovalfiction.com).